Health Anamatics
An Applied Analytics and Informatics Approach Using SAS®

Joseph M. Woodside

support.sas.com/bookstore

The correct bibliographic citation for this manual is as follows: Woodside, Joseph M. 2018. *Health Anamatics: An Applied and Informatics Approach Using SAS®*. Cary, NC: SAS Institute Inc.

Health Anamatics: An Applied and Informatics Approach Using SAS®

ISBN 978-1-62960-881-5 (Hardcopy)

SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513-2414.

August 2018

Contents

About This Book

Description

Summary: Health Anamatics is formed from the intersection of health informatics and data analytics. Healthcare systems generate nearly 1/3 of the world's data, and healthcare stakeholders are promised a better world through data analytics by eliminating medical errors, reducing readmissions, providing evidence based care, demonstrating quality outcomes, and adding cost efficiencies among others. Though Healthcare has lagged behind other industries, the turning point is near with an increased focus across the healthcare sector by way of cost pressures, new technologies, population changes, and government initiatives. There is significant demand to take advantage of increasing amounts of data by utilizing analytics for insights and decision making in healthcare.

Purpose: Having conducted several health analytics and informatics related content courses, I have found a need for a comprehensive current text that combines the clinical healthcare informatics concepts with the applied analytics knowledge using SAS which has led to the concept for this book. The textbook content and learning objectives include health informatics and data analytics concepts, along with applied experiential learning exercises and case studies using SAS Enterprise Miner within the healthcare industry setting.

Author These Authors

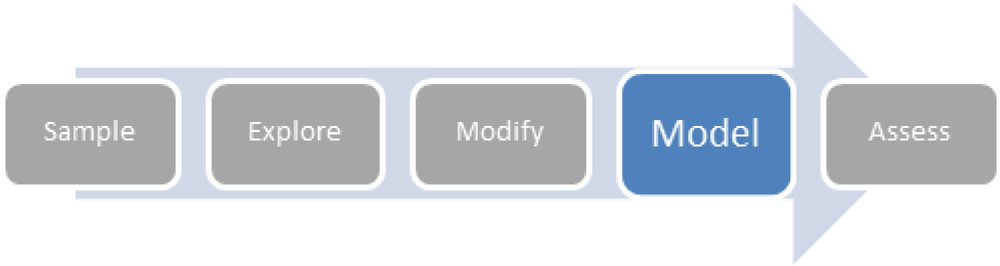
Joseph M. Woodside is an Assistant Professor of Business Intelligence and Analytics at Stetson University teaching undergraduate, graduate, and executive courses on predictive analytics, health informatics, business analysis, and information systems. He has been a SAS user for over nine years, and is responsible for updating the analytics learning goals and course content for the current SAS academic certificate program. Before accepting the Business Intelligence and Analytics position at Stetson, he worked with KePRO, a national health care management company, as the Vice President of Health Intelligence, with responsibility for health care applications, informatics, business intelligence, data analytics, customer relationship management, employee wellness online platforms, cloud-based systems deployment strategy, technology roadmaps, database management systems, multiple contract sites, and program management. He also previously held positions with Kaiser Permanente, with responsibility for HIPAA Electronic Data Interchange (EDI), national Claims and Electronic Health Record implementations, National Provider Identifiers, cost containment financial analytics, and various data analytic initiatives.

Chapter 6: Modeling Payer Data

Chapter Summary

The purpose of this chapter is to develop data modeling skills using SAS Enterprise Miner software, and with respect to the Model capabilities within the SEMMA process. The chapter explores modeling with the decision tree model. This chapter also includes experiential learning application exercises on patient mortality indicators and self-reported general health. The focus of this chapter is detailed in Figure 6.1.

Figure 6.1: Chapter Focus - Model



Chapter Learning Goals

- Describe the model process steps
- Understand payer-level data sources
- Develop data modeling skills
- Apply SAS Enterprise Miner model data functions
- Master the decision tree model

Payers

Payers include insurance agencies and third party payment processors. Typically, payers are separated into categories of commercial payers, governmental payers, and third party payers. The most commonly known type of payer includes commercial payers such as UnitedHealth Group and Blue Cross Blue Shield. Government payers include Medicare and Medicaid which will be covered further in chapter 7. Third party payers include self-insured health plans by an employer, other insurance providers such as health costs covered through care insurance or workers' compensation (PHDSC, 2007). In some cases, providers are setting up independent health plans due to the shift in value-based healthcare to maximize returns. Hospitals could partner with a physician-owned health plan in order to share the cost savings from the value-based care. Hospitals and providers continue to work in hand with insurers to identify high cost or high-risk patients and to improve coordination of care (Livingston, 2016).

Government insurances were originally designed out of a need, due to a lack of coverage available in commercial markets. Medicare was enacted for people over 65, since those over 65 were three times more likely to use medical services, and the costs were unaffordable for both patients and insurers. Medicare is a federal program in the U.S. paid for through payroll taxes. By pooling resources, it enables the protection of individuals in the event of a high-cost healthcare requirement. In addition, there are no exclusions to the program based on age beyond the minimum age of 65, health status or income. Commercial insurance in contrast aims to avoid risk in order to ensure a profit is made and the company remains in business. The for profit aspect allows commercial insurers to exclude those of high risk or high cost, and create barriers to payment of all claims (Archer and Marmor, 2012). TRICARE is another government based U.S. healthcare program for uniformed service members and families and covers general healthcare, prescriptions, and dental plans. The program is managed through the Defense Health Agency and seeks to provide a world-class healthcare system for their over 9.4 million participants. TRICARE offers plans that meet the ACA requirement to maintain minimum coverage (Tricare, 2017). Another related program is the Civilian Health and Medical Program of Veterans Affairs (CHAMPVA), and covers the majority of health expenses. To be eligible for CHAMPVA, members cannot be eligible for TRICARE, and the program has requirements for those with disabilities. CHAMPVA is managed through the Veterans Health Administration, and has over 1,700 locations with over 8.7 million participants. Veterans programs also meet the ACA requirements, and veterans may choose among plans including from the VA, and TRICARE (VFW, 2017; VA.gov, 2017).

Commercial insurance models typically have a model of a Health Maintenance Organization (HMO), Preferred Provider Organization (PPO), or Self-Funded Plan. A HMO model utilizes a primary care physician (PCP) to act as a main point of contact for a patient's care and refers the patients to specialists or other plans of care as appropriate. The PPO model permits patients to see different providers directly and as a result usually carry a higher premium. oth HMO and PPO plans typically have an assigned network of providers with which they contract and negotiate rates. Both HMO and PPO plans also typically pay providers per service rendered or also known as a fee-for-service model (MedicalBillingandCoding, 2017). A self-insured health plan, also known as a self-funded plan, is where the employer pays the financial cost for the healthcare for their employees. With a HMO or PPO plan, a monthly fixed premium is paid for each member, and the insurer pays the financial cost of healthcare. Self-funded plans may represent up to 1/2 of all commercial plans. Employers choose this type of plan to allow more customization of providers and coverage, maintain control, reduce regulations and taxes. Disadvantages include financial risk which may be unpredictable; the self-funded type of plan may be challenging for small businesses or those with a high cost population. Typically, an employer will contract with an existing insurer to administer the self-funded plan, giving a similar network and coverage options with the main difference of financial risk. Over 90% of employers with 5,000 of more workers are self-funded or partially self-funded (SIIA, 2015; KFF, 2016).

In the U.S., the largest commercial health insurers collect over $700 billion in annual premiums, and in 2017 the average annual family premiums were $18,764 (NCSL, 2017). Top insurers in the U.S. include UnitedHealth Group, Kaiser Foundation, WellPoint, Aetna, Humana, Cigna, Highmark, and Blue Cross Blue Shield (BCBS) within various state organizations such as BCBS of California (Heilbronn, 2017; NCSL, 2017). There have been previous attempts between health insurers to pursue mergers and acquisitions within the industry. Anthem offered $48 billion to acquire Cigna, and Aetna sought to acquire Humana for $34 billion. In both cases, federal judges blocked the acquisitions after U.S. Justice Department officials believed the combinations would lead to increased premiums due to reduced competition. Insurance companies argued these would help them negotiate better prices from pharmaceutical companies and hospitals for their customers. The companies are still considering options for appeals, pending potential changes in federal administration, and pursuing various litigation with regard to termination fees and damages due to the failed mergers. With potential changes also planned to the Affordable Care Act and related U.S. healthcare legislation, many insurers are in a wait and see mode, holding on to their cash stockpiles, and determination options for their investment (Tracer et al., 2017; Murphy, 2017). Other healthcare organizations are moving forward, with CVS proposing a nearly $70 billion merger with Aetna (Ramsey, 2018). Cigna acquired Express Scripts, a pharmacy benefit and healthcare management company for $67 billion. Walmart is reportedly reviewing their own options for acquiring Humana health insurance (Pearson, 2018). Walgreens Boots Alliance and wholesale drug distributor AmerisourceBergen have met to discuss a potential $25 billion deal. Other competitors including Amazon, Kroger, and Albertsons have been exploring varying strategies including acquisitions, mergers or alliances with healthcare payers and intermediaries (Hirsch and Sherman, 2018).

Payer Anamatics

Health anamatics has great potential to transform payer cost efficiency and coverage. While analytics and informatics have been utilized by payers for many years for actuarial purposes, usage in other areas of payer operations has varied. Payers are currently focusing on patients, providers, and customers to improve savings. Payers also review financial and operational measures such as forecasting, operations, and fraud monitoring. On the patient side, payers have attempted to identify patients that will have future high costs, in order to increase interventions through preventative care. To assist providers, payers have focused on pay-for-performance programs and fraudulent billing. Similar to the providers, most payers are at stage 2 – localized analytics, of the five levels of analytics capability. While localized analytic capabilities exist, a more complete organizational strategy is still in development, and organizational data warehouse and common analytics toolsets are not prevalent. Payers such as UnitedHealthcare through the acquisition of Ingenix as a subsidiary have developed more advanced analytics capabilities. Payers are also seen as being in an ideal position to utilize health anamatics given the large amounts of claims and transaction data available to them (Cheek, 2014)

Payer Data

Payer Systems and Sources of Data

Payers also generate and collect large amounts of data on patients, providers, and outside sources. Common payer systems include claims, population health management, financial and billing. As part of population health management, insurance sponsored or run personal health records and screenings also collect and store information on patients. Following the screening, wellness systems are offered through insurance carriers such as Florida Blue to promote healthier lives through eHealth education and incentives. Insurers also have an incentive to offering these programs as part of their plans aimed at reducing incurred healthcare costs. Insurers also collect personal data such as medical records and health history when reviewing coverage applications and claims. Some insurers may even access social media data to utilize for example while reviewing a claim. While several payer systems are developed in-house, software vendors of provider systems also develop payer systems. Epic, a company known for their EHR software, has a product called Tapestry for managing health insurance. Features include enrollment verification, member portals, care management to improve health outcomes, customer relationship management module, utilization management, and claims adjudication, processing and billing (Epic, 2017).

Another well-known company for claims processing is TriZetto, which permits electronic claims processing. Trizetto has software edits to reduce errors and improve payments, creates secondary billing for claims with additional insurance payers, send electronic bill remittances, and convert image files to HIPPA compliant EDI formats. Trizetto has built-in business intelligence and analytics for reports and tracking claims data for improving decision making. In 2014, Cognizant announced a $2.7 billion acquisition of privately-held Trizetto. The combined company includes 350 health payers and 180 million covered lives within the U.S. (Cognizant, 2014; Trizetto, 2017). McKesson also offers supporting software for payers such as ClaimsXten, an auditing software to improve accuracy of payments, increase auto-adjudication, convert between ICD 9 and 10 codes, reduce administrative costs, such as through identification of waste and abuse items, and local coverage decisions (McKesson, 2017).

Payer Systems Process

Health claim processing varies by vendor and plan, with advantages for each method. Overall a common system framework exists despite the individual process differences. The first layer includes the data storage, contained within a database management system. The database stores the claims data, provider data, and member data. The system should contain a set of reporting tools for making business decisions, producing regulatory reporting, and providing customers with information. A system should contain a processing engine for develop rules to adjudicate or process claims, including automatic adjudication. The system should also provide a method for customization and modifications of benefits and rules, this is due to the increasing complexity of health claims processing (TM Floyd, 2006).

A typical process would begin first with the completion of a medical treatment. The medical treatment would then be followed by a claim submission to the insurer. The claim form may be mailed on paper or sent electronically. The claim form may be scanned and read or manually entered. The health plan will then review the claim to determine whether payment should be made. The payment determination and in what amount, is the process of adjudication. During adjudication, the health plan will also check the patient benefits and eligibility for services. The provider will also be verified for processing and payment terms. Further checks for other insurance coverage will be made and various quality checks conducted, such as a duplicate claims check. Health plans may also reject the claim or deny the claim without payment. Once adjudication is complete and payment determinations made, an Explanation of Benefits (EOB) will be generated and sent to the insured, which is typically the patient (TM Floyd, 2006).

Claim Forms

Two primary paper claim forms are utilized for billing payers, the CMS-1500 previously known as a HCFA (Healthcare Financing Administration) form for outpatient claims, and a CMS-1450 previously known as a UB-92 (uniform billing) form for inpatient claims. Electronic claims use Electronic Data Interchange (EDI) transactions that are standardized through HIPAA. The HIPAA 837 is the equivalent EDI transaction for the CMS-1500 and CMS-1450 paper forms (CMS, 2014; CMS, 2016). EDI transactions for claims were modeled after the paper forms with additional capabilities. Within the forms and EDI transactions there are a number of common insurance billing terms. Table 1 includes a brief summary table of key terms.

Table 1: Claims Terminology

Term(s)	Definition/Example
Guarantor, Health Plan, Payer	This is the financially responsible party for the claim, such as Aetna.
Subscriber, insured party, enrollee, member, beneficiary	This is the patient that represents the claim. Historically this may have included the parent as the subscriber, however most health plans now bill using the patient information only.
Member number, policy number, insurance ID	This is the unique identifier for the patient. Historically this may have been a social-security-number, however most health plans now assign each patient a unique identifier that does not identify the individual patient.
Group number	This is the unique identifier of the coverage group. Typically, this number is the same for all employees of a given employer, and identifies the coverage and benefits.
Adjudication	This is the process of reviewing and paying the claim by the insurer. The claim may be automatically adjudicated

	or require a human review to determine coverage and payment.
Explanation of benefits (EOB), remittance advice	The explanation of benefits is provided to the patient following insurer adjudication. This form explains the charges from the provider and what was covered or paid by the insurer after adjudication.
Billed amount	This is the original amount billed by the provider for the service(s) rendered, and may include one or more services, such as an office visit, medical supplies, etc.
Allowed amount	This is the maximum amount permitted for the service(s) as per the contract between the insurer and provider.
Contractual Adjustment	This is the difference between the billed amount and the allowed amount per the contract between the insurer and provider.
Coordination of benefits, crossover or piggyback claims	This may occur if a patient has more than one coverage or priority coverage, for example a claim may be first paid by Workman's Compensation as the primary payer if the employee was injured at work, then any remaining amounts may be covered by the employee's commercial insurance as the secondary payer.
Copay, coinsurance, deductible, out of pocket amount	These are the amounts the patient is responsible, even if the claim is paid by the insurer. These help offset the monthly premium costs and are referred to as shared patient responsibilities.

To learn more the claims process and claim forms, we will review claim billing and payment processing examples using the paper-based forms.

Experiential Learning Activity - Claim Forms Billing

Claim Forms Billing

Description: CMS-1500 is the most commonly used claim form, and is used for professional claims such as physician office visits. Navigate to CMS.gov and search for 'CMS-1500', open the CMS-1500 form. For this activity, you will practice completing the paper based forms.

The form is completed through a series of numbered boxes starting with 1-33, some boxes have sub-parts such as 1a. The form completion order follows like reading a page, left to right, top to bottom. Complete the form using sample data for yourself as the patient. Some boxes will require a code lookup, in box 21, you will need to enter the ICD-10 diagnosis code. In box 21 d., you will need to enter the CPT/HCPCS code. There is a more detailed CMS-1500 form instruction available for field by field assistance.

CMS-1500 Form

https://www.cms.gov/Medicare/CMS-Forms/CMS-Forms/Downloads/CMS1500.pdf

CMS-1500 Form Instructions

http://www.cms.gov/Regulations-and-Guidance/Guidance/Manuals/Downloads/clm104c26.pdf

ICD-10 Lookup

https://www.cms.gov/Medicare/Coding/ICD10/2018-ICD-10-PCS-and-GEMs.html

http://www.icd10data.com

HCPCS Fee Lookup

http://www.cms.gov/apps/physician-fee-schedule/overview.aspx

Experiential Learning Activity: Claims Adjudication Processing

Claims Adjudication Processing
Description: For this experiential learning activity, you will take the role of a claims adjudicator. The claim is adjudicated following payer receipt of the paper claim form from the provider. In order to adjudicate the claim properly, coverage information has been provided. Calculate the amounts following review.
Reimbursement Key Terms Summary

Eligibility: Span of insurance coverage based on service date.
Premium: Monthly payment for insurance.
Copay: Fixed amount due at time of each service regardless of deductible/OOP, may vary by type or location.
Deductible: Direct amount of payment before insurance begins.
Coinsurance: Percentage of allowed amount payment after deductible is met.
Out-of-Pocket Maximum: The highest cumulative payment in a calendar year, generally deductible, co-pay and co-insurance all count towards your out-of-pocket maximum.
Contracted Payment: Amount your insurance has agreed to pay the provider of care.

Claim Adjudication

Benefit Coverage Details:
Eligibility Dates: 1/1/2018 – 12/31/2018
Plan: Florida BlueCare Everyday Health
Premium: $325 Per Month
Allowed Amount - Contracted Payment Rate: 80% of Billed
Copay: $0 Preventative (e.g. immunization) / $20 Primary Care Provider / $35 Specialist / $75 ER
Deductible: $500 Per Person / $1,600 Per Family
Coinsurance: 10% of the Allowed Amount (after deductible is met)
Out-of-Pocket Maximum: $2,500 Per Person / $5,000 Per Family

Claim Adjudication

2018 Beneficiary Records (Totals YTD):
Member 11111111:

- Deductible Individual / Family: $0 / $0
- OOP Individual / Family: $0 / $0

Member 22222222:

- Deductible Individual / Family: $500 / $1600
- OOP Individual / Family: $2500 / $5000

Member 33333333:

- Deductible Individual / Family: $100 / $100
- OOP Individual / Family: $20 / $20

Claim Adjudication

Claim # 1000

Member	11111111
Service Date	3/1/2018
Service/Procedure	Office Visit
Billed Amount	$100
Allowed Amount	
Not Covered Amount	
Copay Amount	
Deductible Amount	

Coinsurance Amount	
Patient Responsible	
Insurance Responsible	

Claim # 1001

Member	22222222
Service Date	12/5/2017
Service/Procedure	Office X-Ray
Billed Amount	$150
Allowed Amount	
Not Covered Amount	
Copay Amount	
Deductible Amount	
Coinsurance Amount	
Patient Responsible	
Insurance Responsible	

Claim # 1002

Member	33333333
Service Date	5/15/2018
Service/Procedure	Specialist Visit
Billed Amount	$250
Allowed Amount	
Not Covered Amount	
Copay Amount	
Deductible Amount	
Coinsurance Amount	
Patient Responsible	
Insurance Responsible	

Electronic Data Interchange

Now that you are familiar with the paper-based forms for claims billing, we will cover the alternative electronic method known as EDI, which captures a similar set of data as the paper form and includes the ability to send additional information not included in the paper forms. Electronic Data Interchange (EDI) is the computer to computer exchange of information using international standards.Large retailers such as Wal-Mart, the automotive industry, and the healthcare industry all use EDI. EDI utilizes computerized technology to exchange data and improve processing efficiencies, delivery times, reliability, and quality over existing methods, and EDI allows for standardized and efficient transmission of data between organizations. EDI is included as part of the Health Insurance Portability and Accountability Act (HIPAA) standards, to facilitate administrative cost savings and efficiencies. HIPAA required the Secretary of Department of Health and Human Services to adopt standards to support the electronic exchange of administrative and financial healthcare transactions primarily between healthcare providers and plans. Transaction standards and specifications were adopted by the secretary to enable health information to be exchanged electronically. Implementation guides for each standard have been produced at the time of adoption, and consistent usage of the standards including loops, segments, and data elements, across all guides is mandatory to support the Secretary's commitment to standardization (Woodside, 2013).

The typical healthcare data process flow involves setting the standard transaction set in batch mode through a file transfer protocol (FTP) or other similar transport method over a Value-Added Network (VAN). The process typically results in a transmission/receipt occurrence once per day. The alternative would be a real-time transmission/receipt resulting in multiple transmission/receipts per day, and would utilize HTTP or a similar protocol as the transport method. The EDI x12 standard is then converted to XML, through a variety of third-party applications or custom-built software. The XML data is then stored in a database typically as a character large object (CLOB). Existing applications that need to interface with the various EDI transactional data such as billing systems, claims systems, membership systems, authorization systems, and financial systems, typically cannot read EDI or XML. This results in a secondary conversion to a fixed file format readable by the source system. The data is then stored within the source system for use, resulting in data redundancy within internal systems. For EDI transactions responses, such as a 271, 277, 278, and 835, the process repeats. The source system produces a file in a fixed file format. The file is then converted to XML, which is then converted to the EDI x12 standard (Woodside, 2013).

A clearinghouse, which is an intermediary between providers and health plans, may be used. The typical role of a clearinghouse is to receive EDI transactions from the provider and payer, convert them to the appropriate format and send them on to the appropriate party. A clearinghouse may take non-HIPAA formatted data, and translate to the standard HIPAA EDI format. A clearinghouse may also run quality or edit checks and analytics on the transactions. Typically, a per transaction fee is assessed by the clearinghouse. A clearinghouse is permitted to transmit PHI as they are considered one of the covered entities under HIPAA. Change Healthcare is one of the largest clearinghouses in the U.S. with over 2,100 payer connections, 5,500 hospitals, and 800,000 physicians. In the latest fiscal year, Change Healthcare processed over 12 billion healthcare transactions and $2 trillion in claims (Change Healthcare, 2017).

In effort to reduce the costs of healthcare, which in the U.S. has averaged double the inflation rate per year since the 1970s, EDI standards were created as part of the 1996 HIPAA act. The U.S. is not alone in these efforts: China has implemented measures to promote EDI, including policies, and infrastructure investment. Problems confronting healthcare organizations include increasing costs and inefficiencies in resources. Hospitals began using EDI to communicate with other hospitals, suppliers, insurance companies and banks. The relatively limited EDI presence is explained by high EDI start-up costs as compared with labor, unfamiliar new relationship making, and technical infrastructure and complexity. A New Jersey state study, the HINT project estimated the cost savings from application of computerized systems. Their findings included estimates that 17% of costs are related to processing, and a minimal reduction in those costs would amount to several billion dollars across the industry. Most payers have already put significant investment into computer technology, and can further tap into EDI. One of the most detailed and comprehensive analysis for EDI standards was created by Workgroup for Electronic Data Interchange (WEDI). A large number of estimates were provided, and included pilot projects. WEDI mentioned that although estimated savings may not result in hard-dollar savings, it will allow for efficiency to be improved and resources to be re-allocated to improve quality, care, and service. Additional studies list benefits which include near-term reduction of paperwork, and a long-term potential to use information technology to improve quality and cost effectiveness of healthcare. System data standards integrated across parties will allow for improved accuracy, reliability, and data usage (Woodside, 2007).

As part of the HIPAA legislation, a set of approved EDI transactions were developed to simplify processing and reduce costs. The transactions were developed in compliance with ANSI standards and some documentation includes the ANSI prefix. EDI is popular across many different industries such as finance and manufacturing with different transaction types, and was applied similarly to healthcare. The standard set includes:

Table 2: EDI Transactions

Category	Transaction	Description
Authorization	278	Referral certification and authorization. This is used to request prior authorization for a service and provider referral to ensure payment. Precertification or preauthorization is the prior approval by the payer of a certain action to be taken by the provider during treatment. The claim may be denied if authorization is not requested prior to the service. EDI reduces time spent by the payer contacting one another or the provider. Additional time is reduced by documenting and/or entering data received manually. Assuming 30% referrals, admissions and emergency room visits require review/approval; the payer savings from using the EDI 278 transaction is \$0.81 to \$1.23 per transaction, with provider savings of \$0.65 to \$0.98 per transaction.
Claims	837	Claims or equivalent encounters and coordination of benefits. The 837 is used for billing the claim, similar to the paper form CMS-1450 and CMS-1500. The 837 may have a sub-designation following such as 837-I or 837-P, this designated institutional or professional to match with the same billing of CMS-1450 (institutional claims) and CMS-1500 (professional claims). Claims transactions are simplified through EDI. Information can be entered and transmitted electronically from the

		provider to the payer. Claim information can be re-sent easily, which include claims corrections and adjustments. The estimate of payer savings ranges from \$0.50 to \$1.50, minus a transaction cost of \$0.17. The provider cost per transaction for physician claims varies from \$0.51 to \$1.96, with hospital claims from \$0.11 to \$1.07. Coordination of benefits transactions enables electronic transmission on a single claim. The cost savings potential for payers is \$0.22 per transaction. The savings for providers is \$0.95 to \$1.16 per transaction, based on savings by not identifying, copying, and re-submitting remittances from one payer to another.
Claims – Additional Information	275	Patient Information in Support of a Health Claim or Encounter. The 275 is used for attaching electronic information such as clinical information, lab reports, emergency department, rehabilitative, ambulance services, and medications. This is for supplemental information not included on the 837 EDI transaction.
Claims - Status	276-277	Claim status inquiry (276) and response (277). The 276-277 is a paired transaction to check on the payment status of a claim. The 276 is sent by the provider to the insurer, and the insurer returns a 277 with the current claim status. In the past, providers may have had to call and wait on hold to check the status of claims, now they can have real-time updates as needed. Claims status transactions typically are received by mail or phone. It is estimated that public and private healthcare payers receive over 60 million claim status inquiries per year, and EDI is estimated to save payers \$1.06 to \$2.72 net per inquiry, and save providers \$3.56 to \$3.88 per inquiry.
Claims - Response	835	Remittance and payment advice. Also known as an Electronic Remittance Advice (ERA). The 835 is used to provide the explanation of benefits and payment and describe how the claim was adjudicated and provides details of the claim payments. Payment and remittance transactions include transfer of funds typically by check, and the explanation of the benefit payments from the payer. Potential savings

		include electronic remittance and electronic funds transfer transaction. The savings result from the elimination of postage and handling. The manual costs of processing a remittance and payment range from $0.45 to $1.00, while the costs of processing under EDI are $0.11 to $0.35. The net savings range between $0.10 to $0.89. Approximately $73,432 can be saved per year per hospital, and $1,918 in savings per year per physician practice.
Membership - Eligibility	270-271	Eligibility benefit inquiry (270) and response (271). The 270-271 is a paired transaction to check on the eligibility and benefits of the patient. The 270 is sent by the provider to the insurer, and the insurer returns a 271 with the current claim status. In the past, providers may have had to call and wait on hold to check the status of eligibility, now they can have real-time updates as needed. A patient may not know their copay amount and a provider can easily verify the amount. Eligibility transactions allow confirmation of an individual's eligibility for healthcare services payment by a third party, as well as determining benefit coverage including patient liabilities. An estimated 150 million transactions occur each year, primarily by telephone. The savings estimated for payers is of $0.50-1.00 per inquiry. The savings estimated for providers is $1.10 to $2.09 per inquiry.
Membership - Payment	820	Health plan premium payments. The 820 is used to make monthly payments for the insurance enrollment, typically the set of employees.
Membership - Enrollment	834	Enrollment and de-enrollment in a health plan. The 834 is to add or remove monthly membership in the insurance, typically as employees are hired and leave each month.
Pharmacy	NCPDP 5.1	Retail drug claims, coordination of drug benefits and eligibility inquiry. NCPDP is used for billing pharmacy services, such as at Walgreens. Electronic prescribing (e-prescribing) savings are estimated at $27 billion per year in the U.S. Savings are due to reduced errors, improved efficiency, and easier access to payer drug formularies or approved drugs when

		prescribing, also the ability to substitute lower cost generic drugs or formulary options when available (Porterfield et al., 2014).

EDI Structure

To review the EDI transaction structure, first picture a blank text file. Then to complete a transaction, within the text file you may have different data elements such as the patient name or the amount billed. If everyone completed the text file however they thought best, you would wind up with many different variations of the text file. EDI creates a set of standards and exact positions within the text file to place your data elements such as patient name and amount billed. With this solution, everyone sends and receives their text files the same way and they can be easily translated using EDI.

To begin with basic EDI terminology, there are a series of loops within an EDI text file. Think of these as headings when writing a paper: you have your title, introduction, analysis, and conclusion. Similarly, with EDI loops you have transaction file title information, submission information, patient information, claim information, and individual service information. The loops come into play because you can repeat information at each of the levels. A submitter may submit 100 claims though their information is only needed once. Likewise, a patient may have multiple services and their information is only needed once. A listing of common loops are below, each also a standard alphanumeric number assigned, beginning with the 1000A loop for submitter, and continuing through 2400 for services.

Within each loop there are a series of segments, and each segment has a name and designation. Within the 1000A submitter loop, there is a NM1 segment for name, and a PER segment for contact info. Next, within each segment there are a series of elements, which are numbered with the segment plus 01, 02, 03, etc. Some segments may have more or fewer elements. The NM1 segment has 9 elements, numbered from NM101, NM102, NM103, NM104, NM105, NM106, NM107, NM108, and NM109. In many files, a comma is used to separate values, however a comma may be used in someone's name such as Name, Jr. Using the comma would create an issue when separating a file into elements. Instead, a unique delimiter is used to separate the fields using an asterisk. The special character of tilde ~ is used to end a row or segment. A segment may therefore, look as follows, with each element listed below:

NM1*NM101*NM102*NM103*NM104*NM105*NM106*NM107*NM108*NM109~

Below is another example of a completed segment with Florida hospital included as the provider.

NM1*85*2*FLORIDA HOSPITAL*****XX*1033239991~
N3*3565 S. MAGNOLIA AVE.~
N4*ORLANDO*FL*32806~

Note that some elements contain special codes to designate the following field. In position NM103 there is a '2', which according to the EDI standard designates that this is an organization, while a code of '1' would indicate this is a person such as an individual provider. Note also that there are continued asterisks '*' in sequence, which indicates there are no values. In a segment, there are also required and optional elements. The 'XX' in NM108 indicates that the next value is an NPI number, in this case a sample NPI of 103323991. The segment positions are counted by using the asterisks '*' as the delimiters. Start from the left and count the segment and element such as NM101 = '85', NM012 = '2', … NM109 = '1033239991.

The last part would be the tilde '~' to indicate that the segment is complete. The next segment following the billing name would be the address and city, state, zip. Each segment has a designation, in this case N3 for the address, and N4 for the city, state, and zip.

Table 3 contains a summary table of loops and segments found within an 837-P claim transaction. For simplicity, the key loops, segments, and elements are included. Other loops and segments may include default or standard information on each file. Each EDI transaction such as the 837, 835 or 270, has a slightly different set of loops and segments, though following a similar structure.

Table 3: EDI Loops and Segments

Loop	Name	Segment	Elements
1000A	Submitter Name	NM1	NM101-NM109
1000A	Submitter Contact Info	PER	PER01-PER09
1000B	Receiver Name	NM1	NM101-NM109
2010AA	Billing Provider Name	NM1	NM101-NM109
2010AA	Billing Provider Address	N3	N301
2010AA	Billing Provider City/State/Zip	N4	N401-N403
2010AA	Billing Provider ID	REF	REF01-REF02
2000B	Subscriber Info	SBR	SBR01-SBR-09
2010BA	Subscriber Name	NM1	NM101-NM109
2010BA	Subscriber Address	N3	N301
2010BA	Subscriber City/State/Zip	N4	N401-N403
2010BA	Subscriber Demographic Info	DMG	DMG01-DMG03
20101BB	Payer Name	NM1	NM101-NM109
2300	Claim Info	CLM	CLM01-CLM09
2300	Claim ID	REF	REF01-REF02
2300	Health Diagnosis	HI	HI01-HI02
2400	Service Line	LX	LX01
2300	Professional Line	SV1	SV101-SV109
2400	Service Date	DTP	DTP01-DTP03

Experiential Learning Activity: EDI Translation

EDI Translation

Description: For each of the EDI transactions, translate the information and location of the information connecting your knowledge of EDI and claim processing.
EDI
2010AA: NM1*85*1*CARE*SAM****XX*1234567890~ Loop Name: Provider Name: Provider ID Qualifier Value: Provider ID Quality Location: NM1___ Provider ID Value: Provider ID Location: NM1___
2010BA: NM1*IL*1*SMITH*JANE****MI*222334444~ DMG*D8*19431022*F~ Loop Name: Patient Name: Patient ID: Patient DOB: Patient Gender: DMG02 Designation: DMG03 Designation:
What is the NPI number in the EDI line: NM1*85*2*MAYO CLINIC*****XX*1922074434~
What position is the Provider Name located in the EDI line: NM1*85*2*NEMOURS*****XX*1234567890~
Build the EDI segments for the following provider: Name: JOHN HOPKINS EMERGENCY MEDICAL SYSTEM NPI: 1619903622 Address: 5755 CEDAR LN, COLUMBIA, MD 21044-2912

Now that you are familiar with payer system, data, and claim processing, a variety of payer data is captured that can now be used for analysis. We'll continue with our SEMMA modeling process utilizing data available through payer systems, and review the decision tree model.

SEMMA: Model

Model Process Step Overview

During the modeling process step, the data mining model is applied to the data. During the partitioning phase data is segmented into training and validation datasets. The training dataset is used to fit the model, while the validation dataset is used to validate the model on a new set of data to demonstrate the reliability of the model. Based on the results the model can then be tuned to optimal performance.

There are many different models that can be selected during this step. The decision to choose each model is based on earlier exploration of data and knowledge of each model. We will continue this chapter with the model of decision tree.

Model Tab Enterprise Miner Node Descriptions

Model application examples

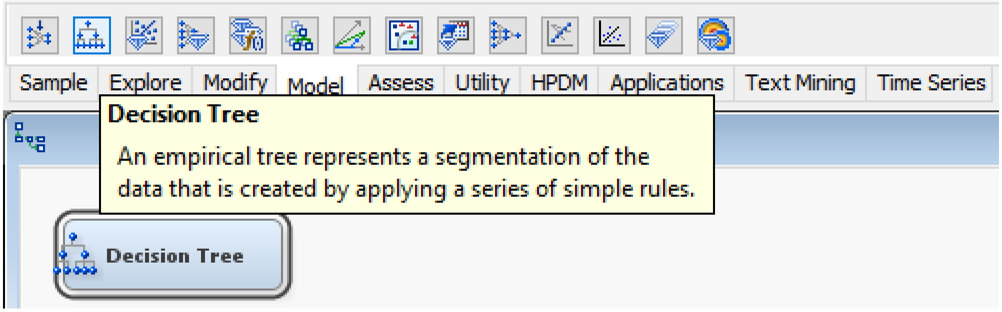
Decision trees can be applied to a variety of areas with the healthcare setting. Researchers in Taiwan examined ICD-9-CM codes within claims data to identify cases of coronary artery bypass graft infections from a sample of 1,017 surgeries. The overall goal of the researchers was to accurately predict infection sites, in an effort to improve quality. Their decision tree model performed well in terms of true positive predictive performance. A set of regression models were also run to compare performance, however researchers noted limitations in the regression model's ability to handle the highly dimensional data, and the decision tree was able to more easily classify highly dimensional data. The first branch or split in the decision tree was length of stay variable (Yu et al., 2014).

Another study, MVP Health Care, with over 750,000 members in the eastern U.S. implemented a set of decision trees for prior claim authorization. MVP Health Care estimated $2 million in savings associated with the improved prior authorizations. Typical prior authorizations previously took 2-3 days' turnaround, cost $75 each and required a phone call. The decision tree made the results available through a web-based interface. The interface walked the provider staff through a few short questions, with the results determined based on the individual patient benefits and medical information. Most medical policies and technology assessments are not standardized and often out of date. The medical policies and guidelines can be updated dynamically, and can be linked to electronic health records to improve information transparency among stakeholders. The policies can be standardized for systematic communication, and centralized for more timely updates (Moeller, 2009).

DecisionTree Node

The decision tree node is utilized for the decision tree model and is found under the Model tab in SAS Enterprise Miner.

Figure 6.2: DecisionTree Node



Decision Tree model description

Decision Trees are a flexible model capable of handling various input and target data types, along with missing and non-standardized data. Decision trees can handle binary, continuous, or nominal inputs and output variables, whereas most models have more specific variable type requirements. Decision trees also do not require the statistical assumptions that must be met with models such as multiple linear regression. As a result, decision trees are one of the most popular and widely used techniques as they can also be easily communicated. Decision trees are modeled after actual trees, though in contrast with living trees, decision trees are often depicted top to bottom or left to right. Just like trees, decision trees are grown starting with the primary node or trunk of the tree and follow a series of branches, segments or splits based on the variables in the dataset. The decision tree is fully grown following a series of splits or branches to the terminal nodes or leaves of the tree (Klimberg and McCullough, 2013).

Model assumptions and data preparation

Due to the overall flexibility, decision trees carry less model assumptions and requirements. Decision trees as with regression may suffer from overfitting the model where the model is perfected for training data and unable to model new data, and again select the parsimonious model or simplest-best model. An advantage of the decision tree is that limited data preparation is required as compared with other models. Decision trees handle missing data and outliers to a greater extent and are less affected than other methods. While data quality is always important, often time constraints impact the model selection.

Partitioning requirements

For decision trees we typically create two datasets: 1. Train and 2. Validate. In some cases, three datasets may be used: 1. Train, 2. Validate and 3. Test. When using the third dataset for the decision tree process, first train the tree or grow the tree to its full potential, then validate or prune the tree to remove extemporaneous or invaluable branches and paths to simplify or improve the shape of the tree, and lastly test the tree using the pruned model from validation.

Model results evaluation

To evaluate the results of our decision tree we can use several items including errors, lift, misclassification rate, and English rules.

Errors

The errors are calculated from the predicted value less the actual value also known as the residuals. Common measures of errors are sum of squared error (SSE) and root mean square error (RMSE). The measures are calculated by taking the square or square root of the errors.

Lift

A measurement between a random or baseline model against the analytical model. A higher lift or outperformance of the random selection is best.

Misclassification Rate

For models with nominal or binary targets, the percentage of total records misclassified as false positive or false negative.

English Rules

For decision trees, a tree model is available which can produce a set of If…Then… conditions also known as English rules which assist with the interpretability of the model. The rules are often used in decision support systems to model human decision making. If you visit your physician they will based on their experience instinctually walk through a set of If..Then… rules to make a diagnosis, for example If you have a fever and cough, and stuffy nose, Then you are diagnosed with influenza or the flu.

Now that we have covered the decision tree model, let's continue with an experiential learning application to connect your knowledge of decision trees with a health application on patient mortality indicators.

Experiential Learning Application: Patient Mortality Indicators

Many quality improvement approaches to improve quality of care are based on manual activities without a direct link to the data within the healthcare information system. Payer and provider systems can supply patient outcome information and clinical pathways to support patient care and factors influence quality of treatment and cost of care. Data mining through decision trees allows for knowledge discovery from large sets of data can be used to identify patterns or rules (Woodside, 2010).

A decision tree can be utilized to determine how inpatient mortality rates compare to overall proportions, and which segments to focus on. In one study, a set of 8,405 patients for indicators of inpatient mortality as part of decision tree analysis to determine inpatient mortality factors. Factors and indicators included gender, discharge location such as surgery department, age group, and disease class. The results found that for patients discharged from Internal medicine departments, mortality was nearly 3 times more likely. Patients with LOS over 16 days, resulted in a 6 times higher mortality rate. The variable significance included LOS, discharge department, followed by age group. While logistic regression could be utilized, the output would be missing the segment characteristics that would be useful (Chae et al., 2003).

Prior studies have examined factors such as gender, discharge department, age group, and disease class to determine if these have a relationship on mortality. For this experiential learning application, we want to verify which of these factors may have a relationship with mortality. To start our process, we first want to identify our input and target variables. In this application our inputs (x) are gender, discharge location, age group, and disease class, and our target (y) is patient mortality. With a decision tree, one advantage is that the model can handle varying input and target variable types. In this application our inputs are nominal and our target variable is binary.

Dataset File: 6_EL1_Patient_Mortality.xlsx

Variables:

- ID, unique identifier
- Gender, (Female, Male)
- Discharge Department, (Internal Medicine, Surgery)
- Age, (Under 20, 21-40, 41-60, 61 or older)
- Disease Class, (Circulatory, Congenital, Eye and Ear, Gastrointestinal, Miscellaneous, Muscle, Neoplasm, Pulmonary, Urinary)
- Length of Stay, (1-5 Days, LOS 17-341 Days, LOS 6-16 Days)
- Inpatient Mortality, (1=True, 0=False)

Step 1: Sign-in to SAS On Demand.

Step 2. Open the SAS Enterprise Miner Application (click on SAS Enterprise Miner link).

Step 3. Create a New Enterprise Miner Project (click New Project…).

Step 4: Use the default SAS Server, click Next.

Step 5: Add Project Name *PatientMortalityIndicators*, click Next.

Step 6: SAS will automatically select your user folder directory (if using desktop version, choose your folder directory), click Next.

Step 7: Create a New Diagram *PatientMortalityIndicators*(Right-click on Diagram).

Step 8: Add a File Import node (click the Sample tool tab, drag node into the diagram workspace).

Step 9: Click on the File Import node, and review the property panel on the bottom left of the screen.

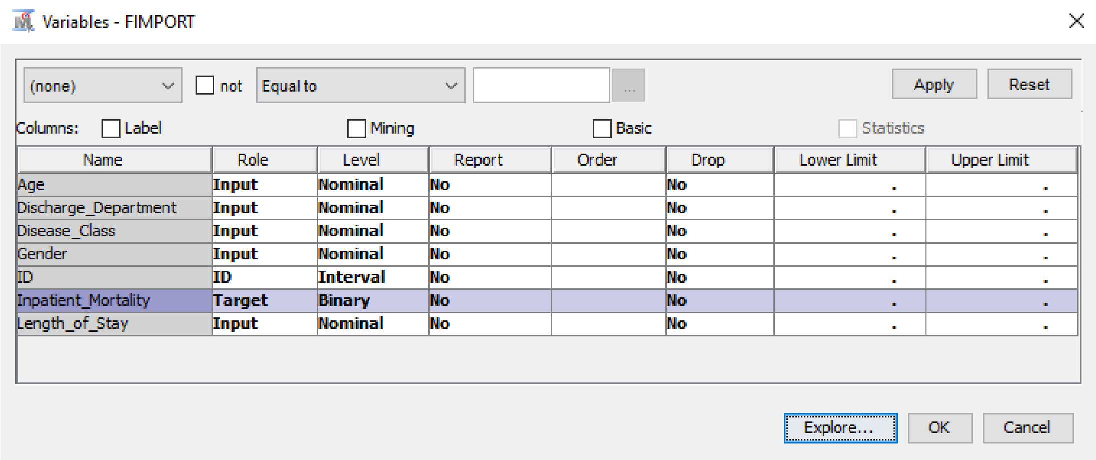
Step 10: Click on the Import File … and Browse to the 6*_EL_1_Patient_Mortality.xlsx* Excel File.

Step 11: Click Preview to ensure the data set was selected successfully, click OK.

Step 12: Right-click on the File Import node and click Edit Variables.

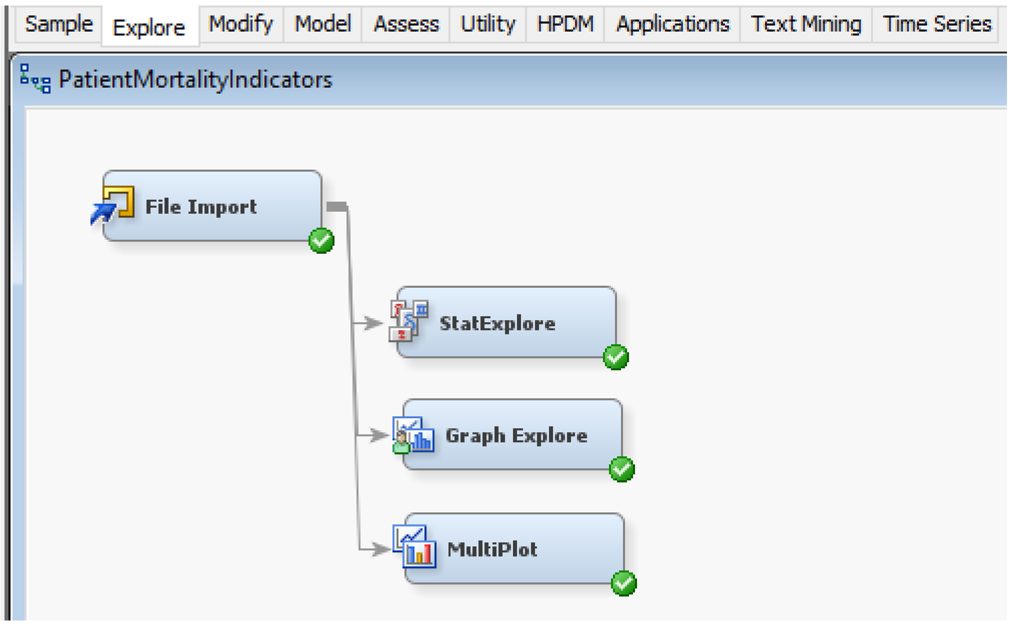
Step 13: Set Inpatient_Mortality to the Target variable role, set ID to the ID role, and all other variables to the Input role. Set the remaining variables according to their nominal, interval, or binary levels. To review an individual variable to verify its role and level assignment, click on the variable name and click Explore. Once complete with setting all variables, click OK.

Figure 6.3 Edit Variables



Step 14: Add a Stat Explore node, Graph Explore node, and MultiPlot node (click the Explore tool tab, drag nodes into the diagram workspace). Set the Graph Explore Property Sample Size to Max.

Figure 6.4 StatExplore and Graph Explore Nodes



Step 15: Review results. From the Stat Explore descriptive statistics results we identify a good data quality result, verified through 0 missing records across variables. The breakout of the target variable Inpatient_Mortality is shown with 8235 records with a value of 0 or False, and 170 records with 1 or True.

Figure 6.5 StatExplore Results

Output

Data Role=TRAIN

Data Role	Variable Name	Role	Number of Levels	Missing	Mode	Mode Percentage	Mode2	Mode2 Percentage
TRAIN	Age	INPUT	4	0	41-60	31.22	Under 20	25.02
TRAIN	Discharge_Department	INPUT	2	0	Internal Medicine	72.71	Surgery	27.29
TRAIN	Disease_Class	INPUT	9	0	Neoplasm	28.76	Miscellaneous	22.39
TRAIN	Gender	INPUT	2	0	Male	52.96	Female	47.04
TRAIN	Length_of_Stay	INPUT	3	0	1-5 Days	52.91	LOS 6-16 Days	31.95
TRAIN	Inpatient_Mortality	TARGET	2	0	0	97.98	1	2.02

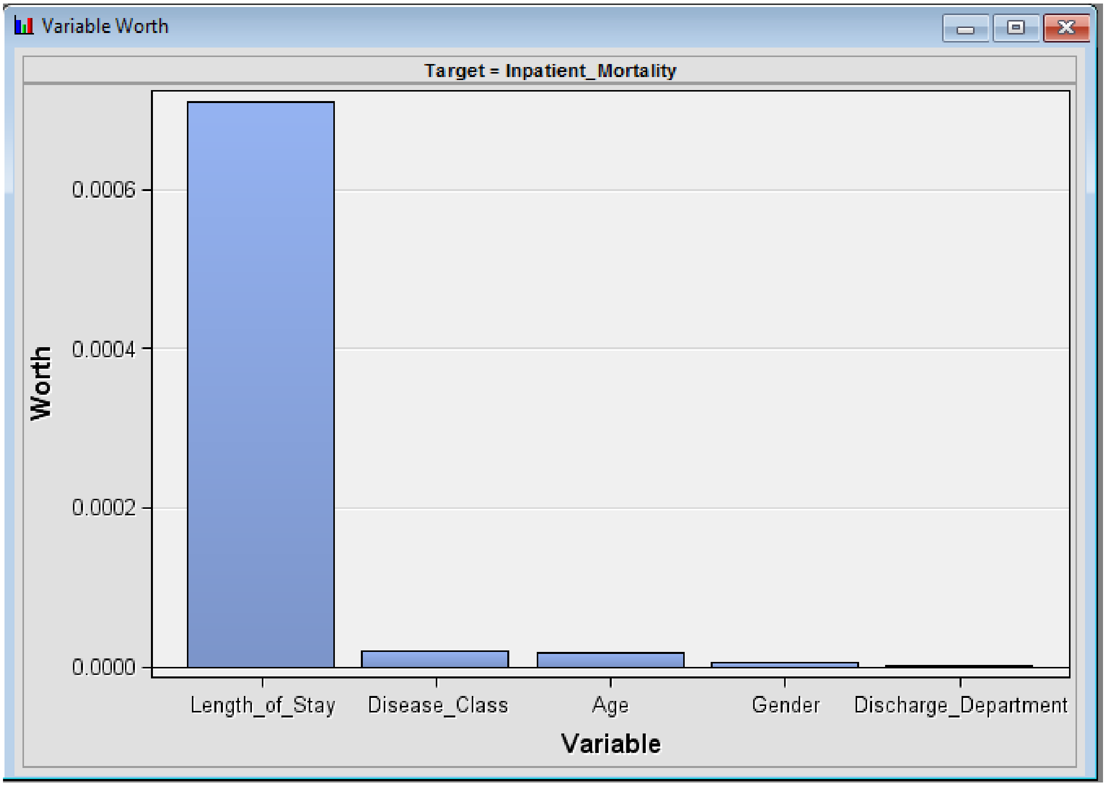
Distribution of Class Target and Segment Variables
(maximum 500 observations printed)

Data Role=TRAIN

Data Role	Variable Name	Role	Level	Frequency Count	Percent
TRAIN	Inpatient_Mortality	TARGET	0	8235	97.9774
TRAIN	Inpatient_Mortality	TARGET	1	170	2.0226

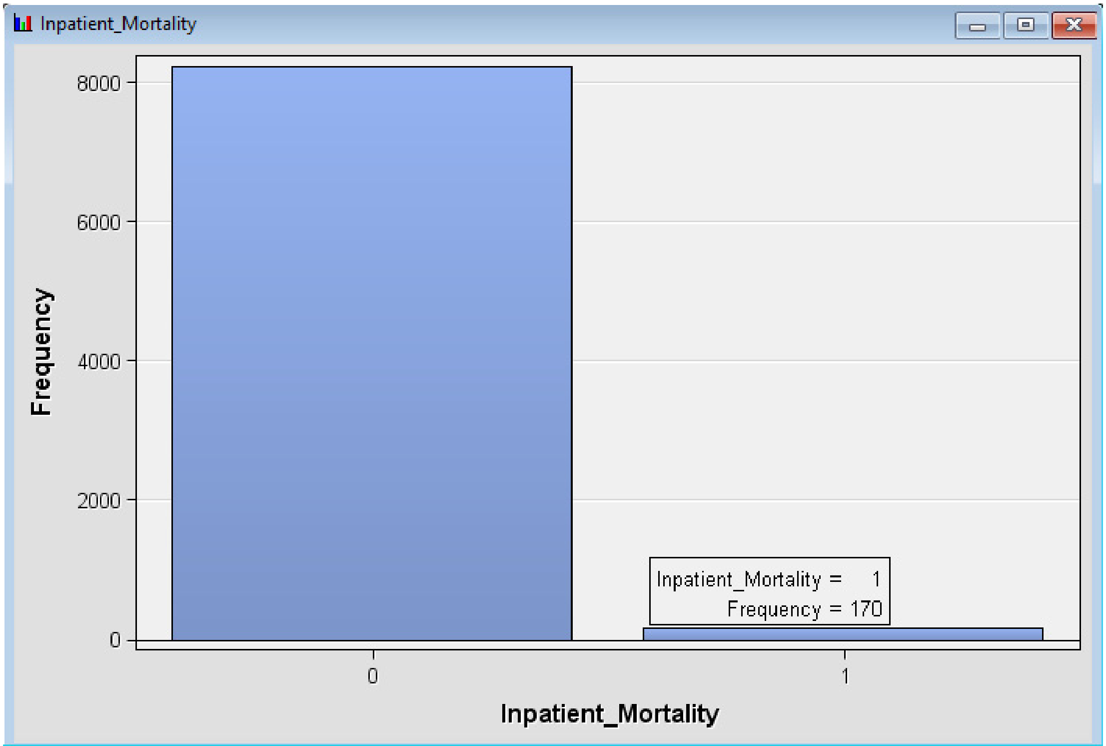
Step 16: Review results. From the Stat Explore variable worth results, Length_of_Stay has the greatest variable worth with regard to our target variable of Inpatient_Mortality

Figure 6.6 StatExplore Results Variable Worth



Step 17: Review results. From the Graph Explore results we also see the breakdown of the Inpatient_Mortality.

Figure 6.7 Graph Explore Results

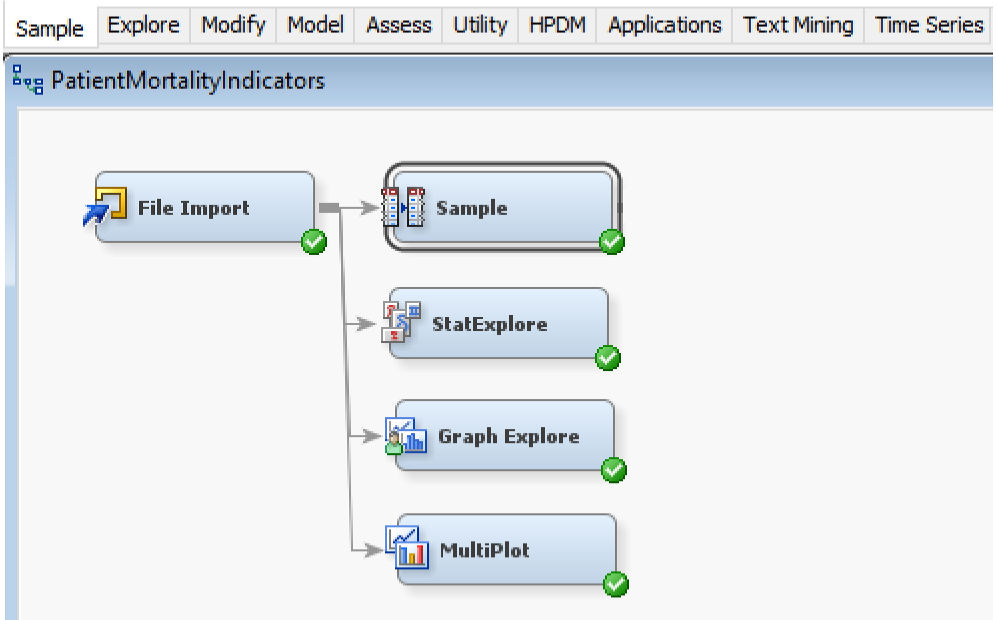


Step 18: Review results. From the MultiPlot results review each of the variables, for example age by inpatient_mortality shows a distribution across all age groups with both 0 and 1 frequencies.

Figure 6.8 MultiPlot Results

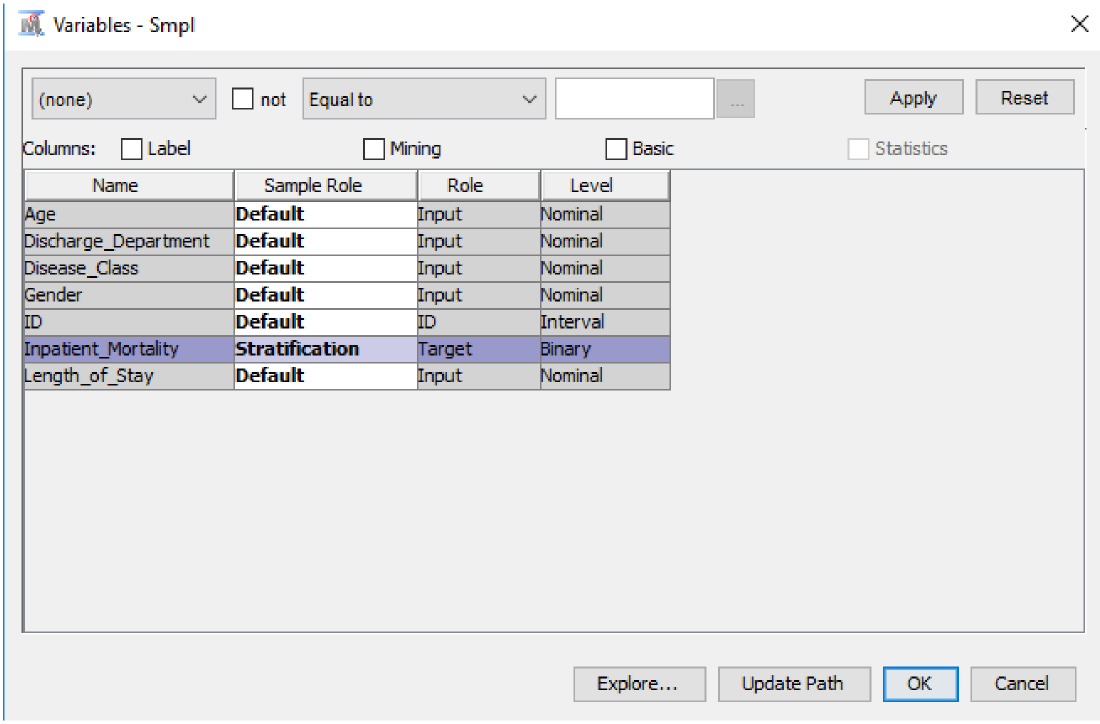
Step 19: From our Stat Explore, Graph Explore, and MultiPlot results, we see that the inpatient mortality is a rare event in terms of occurring only 2% in our dataset. As a result we will include a sampling node to conduct a rare event sampling to improve the final results. From the Sample tab add a Sample node to the diagram and connect to File Import.

Figure 6.9 Add Sample Node



Step 20: Click on the Sample node and click on the properties section, click on Variables… Set the Inpatient_Mortality to Stratification Sample Role. The setting will allow us to select a sample based on the Inpatient_Mortality variable. Click OK.

Figure 6.10 Sample Properties Stratification



Step 21: Click on the Sample node properties can set the Type to Percentage and set the Percentage to 100. For the Stratified property set the Criterion to Equal. The settings will select an equal sample of the Inpatient_Mortality rare event and the Inpatient_Mortality non-event. In other words will select an equal sample of both true and false cases of Inpatient_Mortality. If we selected the normal sample size, the results may be limited given the size of non-events, since all occurrences would favor a non-event scenario. Our goal is to find the factors leading to Inpatient_Mortality.

Figure 6.11 Sample Properties

Property	Value
Size	
Type	Percentage
Observations	.
Percentage	100.0
Alpha	0.01
PValue	0.01
Cluster Method	Random
Stratified	
Criterion	Equal
Ignore Small Strata	No
Minimum Strata Size	5
Level Based Options	
Level Selection	Event
Level Proportion	100.0
Sample Proportion	50.0
Oversampling	
Adjust Frequency	No
Based on Count	No
Exclude Missing Levels	No

Step 22: Run the Sample node and view the Results. From the output the original dataset is shown with Inpatient_Mortality = 1/True occurring 170 times for 2% of the total dataset. After sampling, Inpatient_Mortality = 1 occurs the same amount as Inpatient_Mortality = 0 for an equal sample dataset.

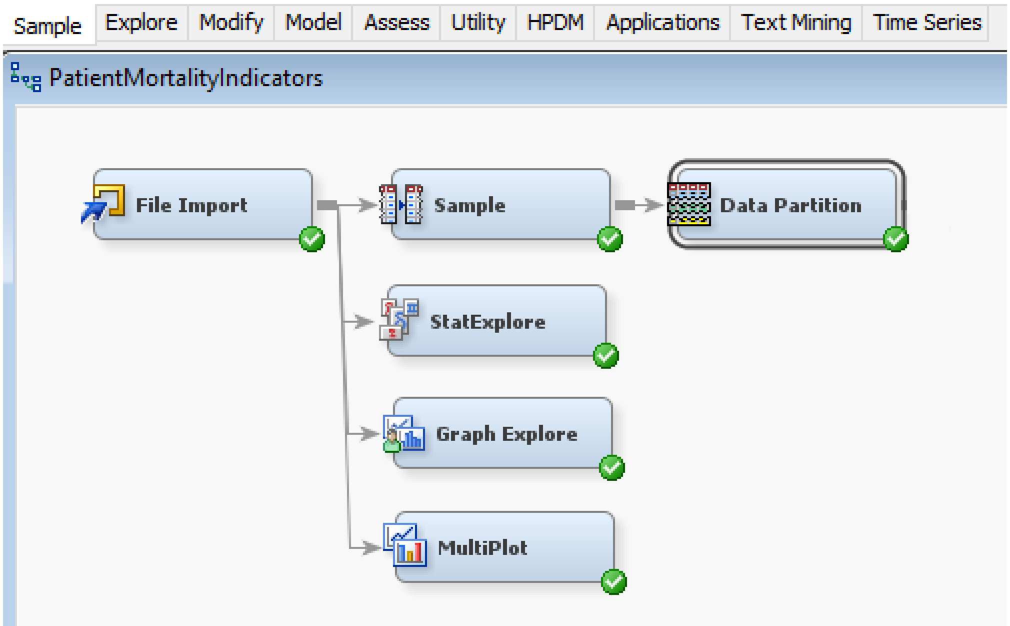
Figure 6.12 Sample Results

Output

```
46
47   Data=DATA
48
49                          Numeric   Formatted   Frequency
50        Variable           Value      Value       Count      Percent         Label
51
52   Inpatient_Mortality       0          0          8235      97.9774    Inpatient Mortality
53   Inpatient_Mortality       1          1           170       2.0226    Inpatient Mortality
54
55
56   Data=SAMPLE
57
58                          Numeric   Formatted   Frequency
59        Variable           Value      Value       Count      Percent         Label
60
61   Inpatient_Mortality       0          0          170         50       Inpatient Mortality
62   Inpatient_Mortality       1          1          170         50       Inpatient Mortality
```

Step 23: Add a Data Partition node (click the Sample tool tab, drag node into the diagram workspace). Set the Data Partition Property Data Set Allocations to 60.0 for Training, 40.0 for Validation, and 0.0 for Test. Run the Data Partition node.

Figure 6.13 Data Partition Node



Step 24: Review the data Partition Results.

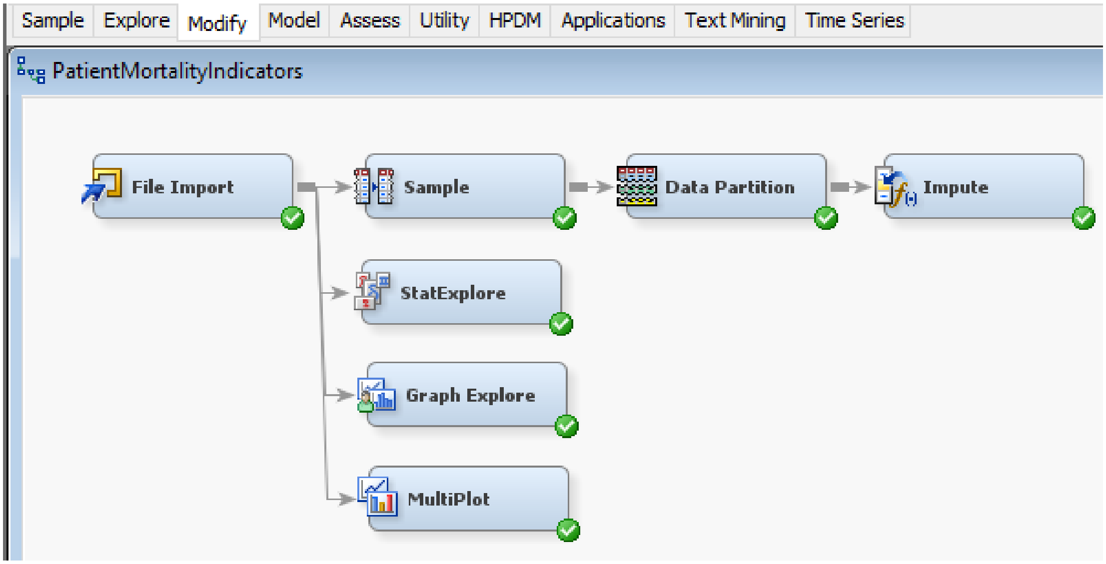
Figure 6.14 Data Partition Results

Output

```
24    Partition Summary
25
26                                          Number of
27    Type            Data Set             Observations
28
29    DATA        EMWS1.Smpl_DATA               340
30    TRAIN       EMWS1.Part_TRAIN              203
31    VALIDATE    EMWS1.Part_VALIDATE           137
32
```

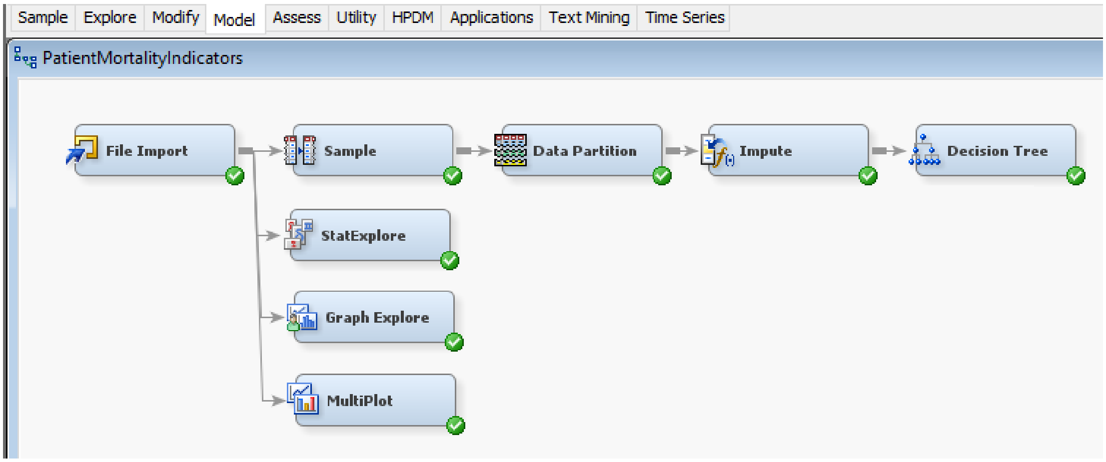
Step 25: Add an Impute node (click the Modify tool tab, drag node into the diagram workspace). Verify the Impute Property is set to Count for Class variables and Mean for Interval variables.

Figure 6.15 Impute Node



Step 26: Add a Decision Tree node (click the Model tool tab, drag node into the diagram workspace).

Figure 6.16 Decision Tree Node



Step 27: Select the Decision Tree node, in the Tree Property under Splitting Rule set Minimum Categorical Size = 2, under Node set Leaf Size = 1. The settings allow a tree to grow with a 2-category split and a single leaf or a single record.

Figure 6.17 Decision Tree Node Properties

Property	Value
Splitting Rule	
Interval Target Criterion	ProbF
Nominal Target Criterion	ProbChisq
Ordinal Target Criterion	Entropy
Significance Level	0.2
Missing Values	Use in search
Use Input Once	No
Maximum Branch	2
Maximum Depth	6
Minimum Categorical Size	2
Node	
Leaf Size	1
Number of Rules	5
Number of Surrogate Rule:	0
Split Size	.

Step 28: Select the Decision Tree node, in the Tree Property select 'Largest Tree'. The setting will run the full tree to all its branches and leaves, or all splits and decision points.

Figure 6.18 Decision Tree Node Properties

Property	Value
Subtree	
Method	Largest
Number of Leaves	1
Assessment Measure	Decision
Assessment Fraction	0.25
Cross Validation	
Perform Cross Validation	No
Number of Subsets	10
Number of Repeats	1
Seed	12345
Observation Based Import	
Observation Based Import	No
Number Single Var Importa	5
P-Value Adjustment	
Bonferroni Adjustment	Yes

Step 29: Right-click on the Decision Tree node and click Run.

Step 30: Expand Output Window Results and Review Model Results. The misclassification rate for the Training set is 31.5% and the misclassification rate for the Validation set is 35.0%.

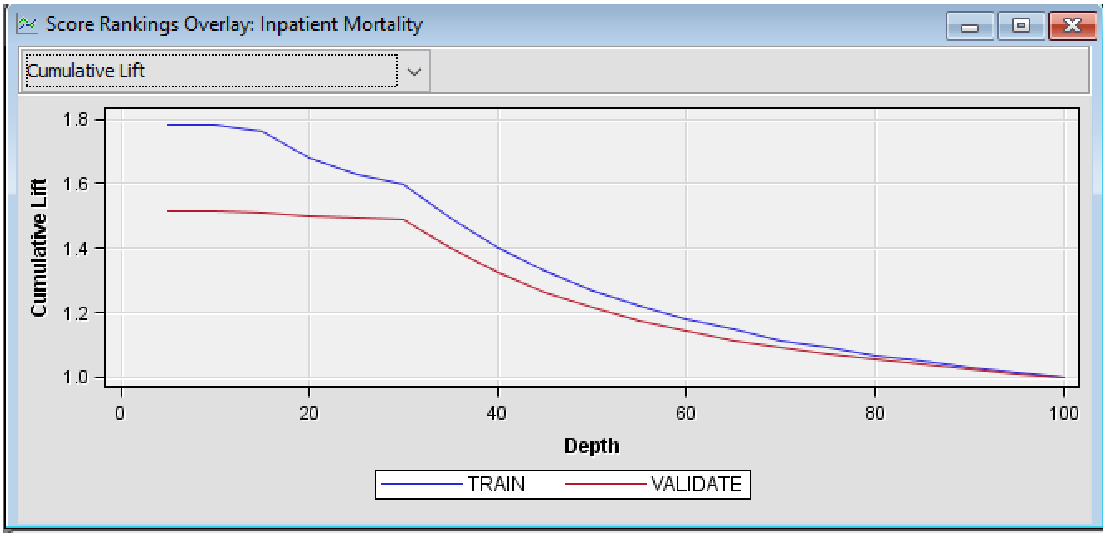
Figure 6.19 Decision Tree Node Results

Fit Statistics

Statistics Label	Train	Validation
Sum of Frequencies	203	137
Misclassification Rate	0.315271	0.350365
Maximum Absolute Error	0.896552	0.896552
Sum of Squared Errors	84.24056	62.3695
Average Squared Error	0.207489	0.227626
Root Average Squared Error	0.45551	0.477102
Divisor for ASE	406	274
Total Degrees of Freedom	203	.

Step 31: Review Model Results. The cumulative lift shows that the model outperforms a random model. At the top 10 percent of records or depth the train model outperforms a random model by nearly 1.8 times and the validation model by 1.5 times.

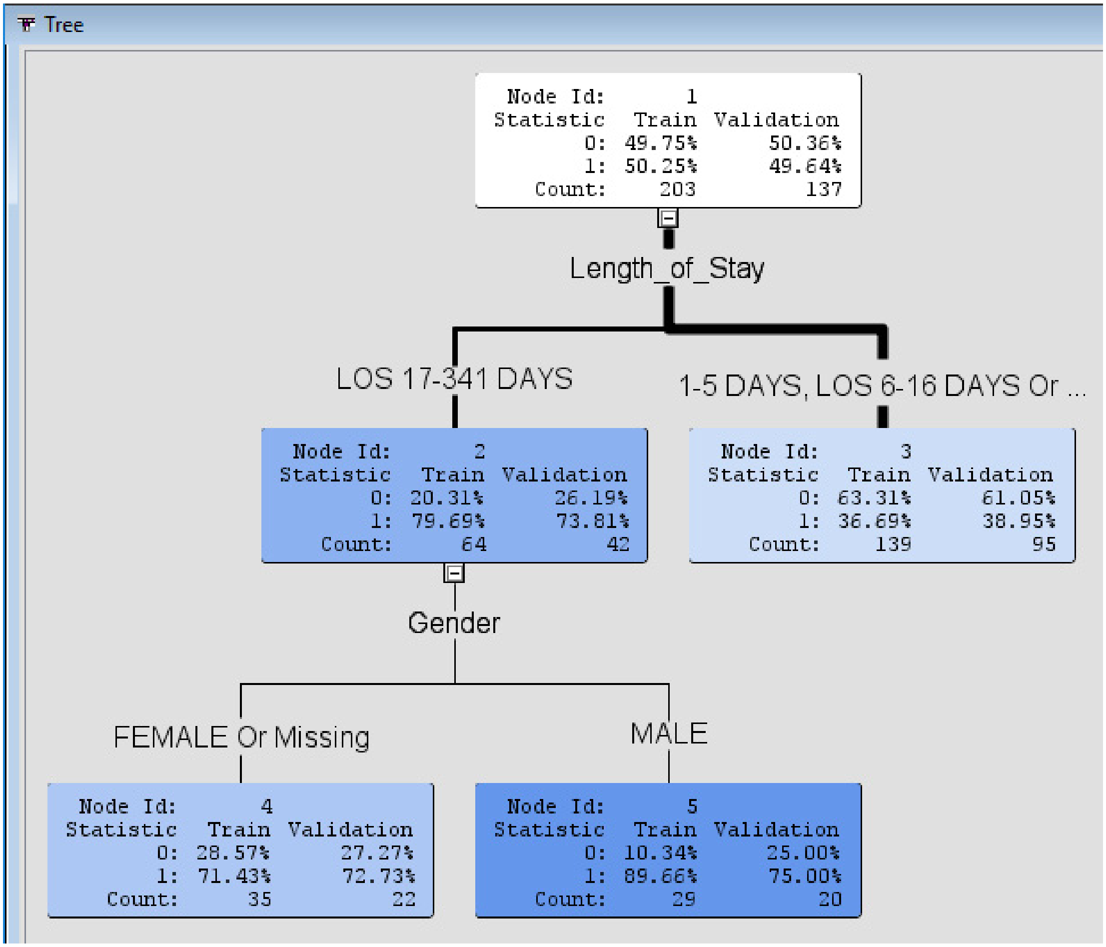
Figure 6.20 Decision Tree Node Results



Step 32: Review Model Results. The decision tree is presented as a visual model. Think of the decision tree modeled after a living tree. At the top we have the root or Node Id 1. The top is where the tree starts like the trunk of a tree. The Node Id 1 also gives a breakdown of our data with roughly 50% 0 and 50% 1 cases, and a split between Train with 203 records and Validation with 137 records. The lines can be considered branches and the remaining nodes leaves, we therefore build or grow our tree starting with the root or Node Id 1 and branching out all the way to the final leaves Node ID 5, also known as terminal nodes. The lines or braches are also different widths or thickness based on the number of records. Try to visualize an upside-down tree in the Figure 6.23. Start again from the root Node Id 1, the first split in the

tree occurs with the variable Length_of_Stay, this indicates that Length_of_Stay has a high variable importance in our model. If we follow the split to the left we have LOS 17-341 Days, this means that all records with LOS 17-341 Days follow the left side of the split. The right side of the Length_of_Stay split contains all records with LOS 1-5 Days, 6-16 Days or missing values. Looking closer at Node Id 2 the record breakdown is also given. For Validation, 26.19% are 0 cases and 73.81% are 1 cases, this indicates that using a LOS of 17-341 Days split, we can identify 80% of the 1 cases or true cases for patient mortality. We can further follow the tree to the next split and branches of Gender. Follow the tree to Gender of Male and Node Id 5. For the Validation breakdown we find 75% are 1 cases and 25% are 0 cases, with true cases slightly higher for males than females.

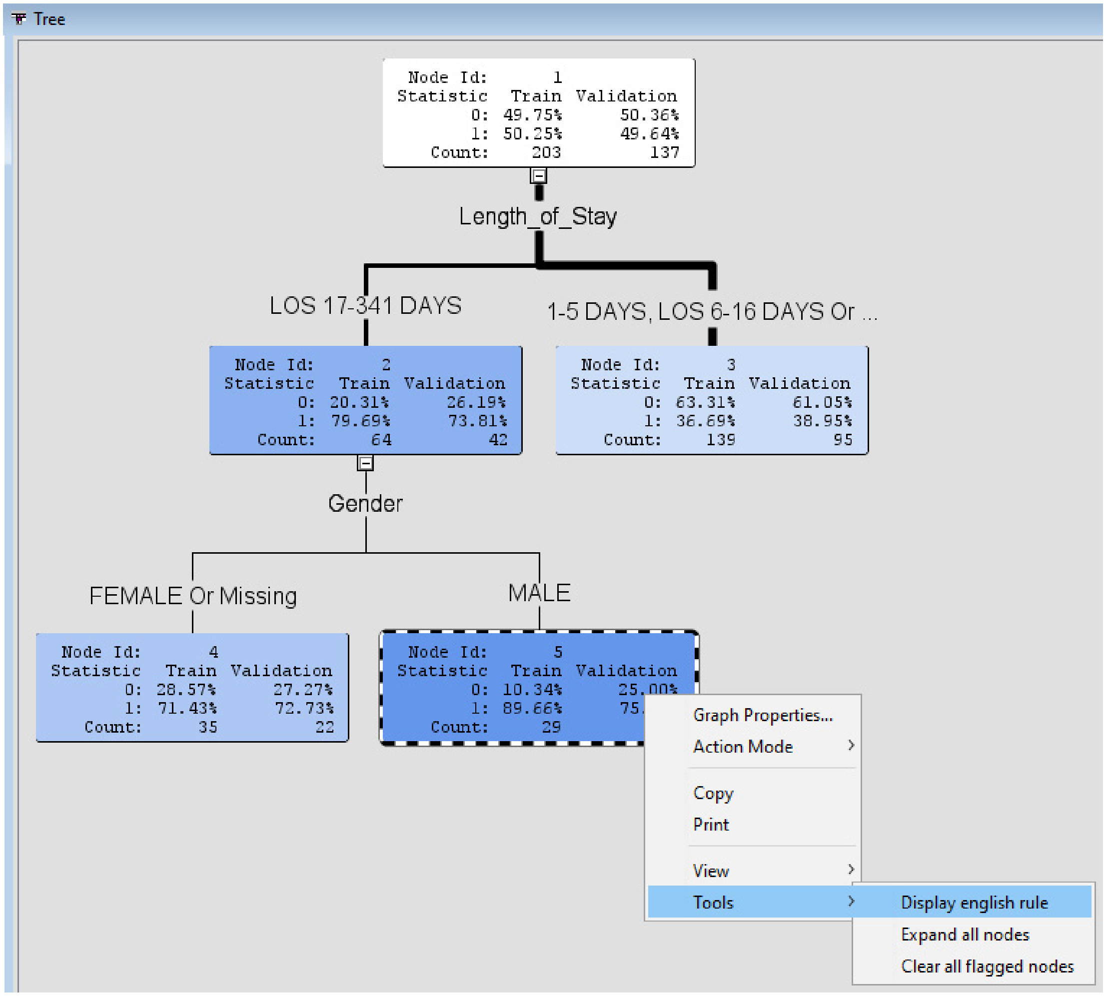
Figure 6.21 Decision Tree Node Results



Step 33: Review Model Results. We can further develop programmatic IF…THEN rules to describe the branches, these are also known as English Rules. The rules can be valuable for automating decision making in a system such as a decision support or EHR system that used by medical professionals for assessing risk

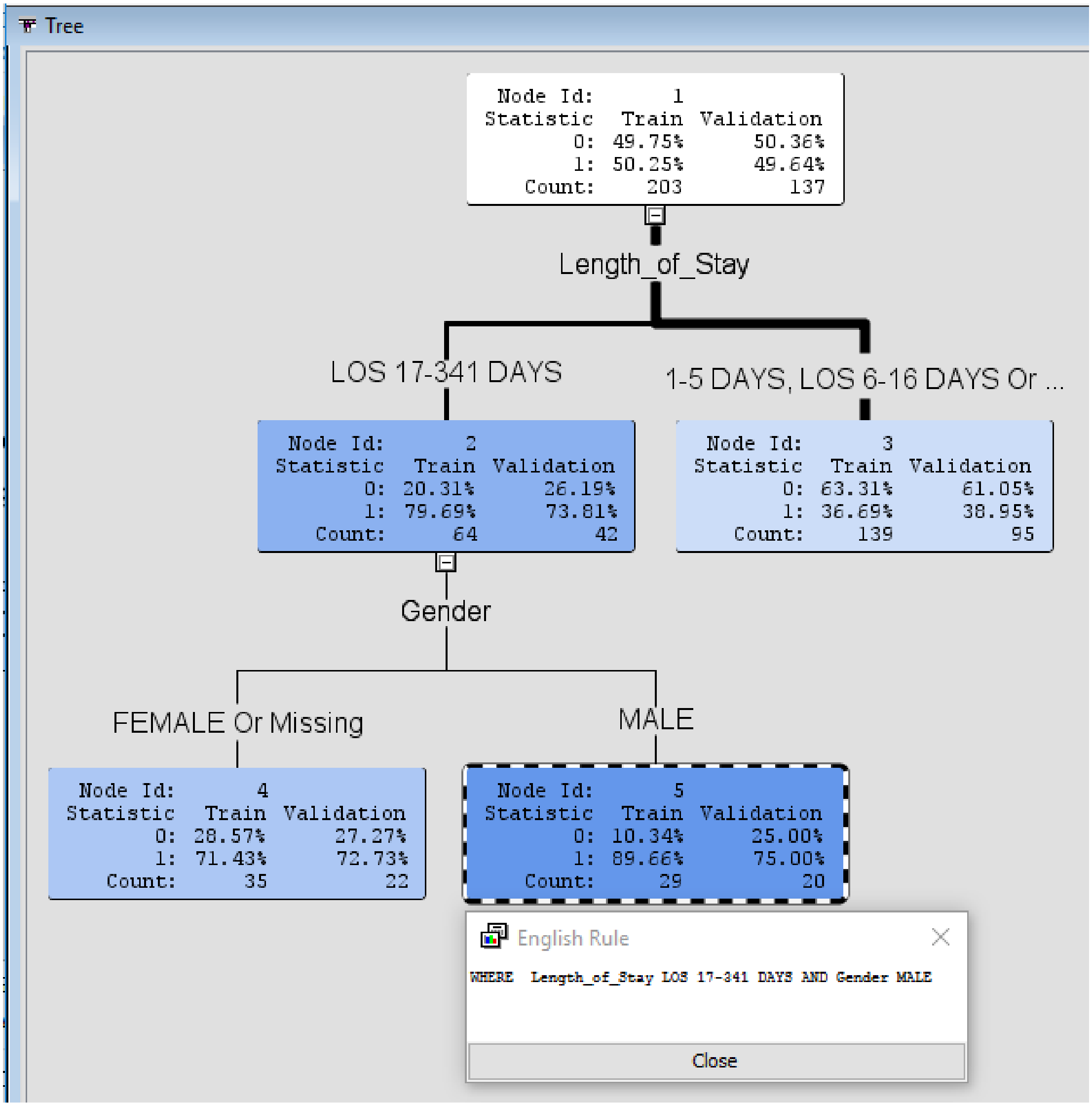
of morality and developing an appropriate plan of care. Click on Node Id 5 and right-click and select Tools → Display english rule.

Figure 6.22 Decision Tree Display English Rule



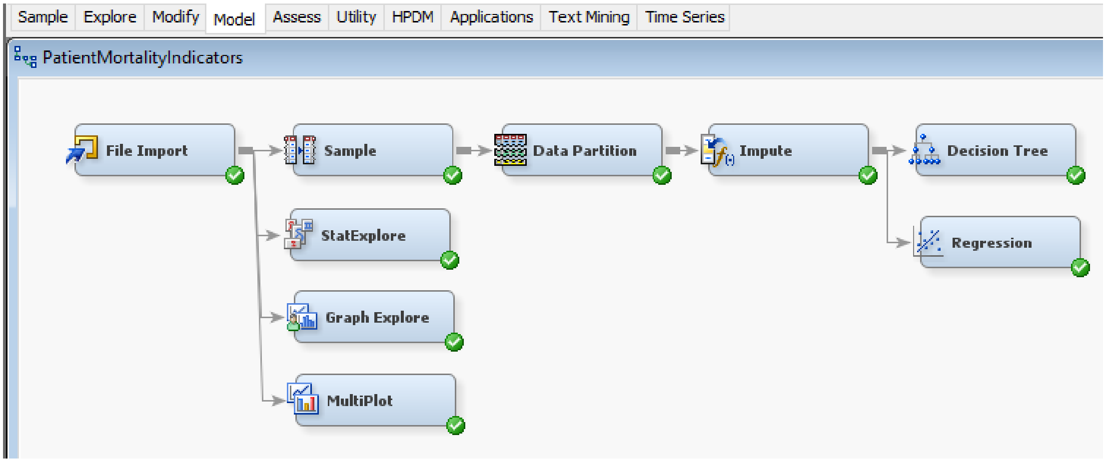
Step 34: Review Model Results. The results will display the English Rule following the tree structure Where Length_of_Stay LOS 17-341 AND Gender MALE. The rules are easily communicated to other clinical and non-clinical individuals, for example rephrasing as if the length of stay is between 17 and 341 days and the patient is male, nearly 90% of our training records and 75% of our validation records indicate patient mortality.

Figure 6.23 Display English Rule



Step 35: Add a Regression node. From our prior chapter we can also include a Regression model for our dataset. The node can be connected similar to the Decision Tree node. Run the Regression node.

Figure 6.24 Regression Node



Step 36: Review Model Results. The misclassification rate for the Training set is 29.1% and the misclassification rate for the Validation set is 38.7%. The results perform similarly to our Decision Tree model with 31.5% and 35% misclassification rate respectively.

Figure 6.25 Regression Node Fit Statistics

Fit Statistics

Statistics Label	Train	Validation
Akaike's Information Criterion	260.2229	.
Average Squared Error	0.191212	0.244899
Average Error Function	0.562125	0.697723
Degrees of Freedom for Error	187	.
Model Degrees of Freedom	16	.
Total Degrees of Freedom	203	.
Divisor for ASE	406	274
Error Function	228.2229	191.1761
Final Prediction Error	0.223933	.
Maximum Absolute Error	0.878243	0.930906
Mean Square Error	0.207573	0.244899
Sum of Frequencies	203	137
Number of Estimate Weights	16	.
Root Average Sum of Squares	0.437278	0.494873
Root Final Prediction Error	0.473216	.
Root Mean Squared Error	0.455601	0.494873
Schwarz's Bayesian Criterion	313.2342	.
Sum of Squared Errors	77.63219	67.1024
Sum of Case Weights Times Freq	406	274
Misclassification Rate	0.29064	0.386861

Step 37: Review Model Results. Adding the Regression node also provides additional results such as Odds Ratio. From the Analysis of Maximum Likelihood Estimates output the Length_of_Stay is a significant variable with Pr > ChiSq, p-value less than 0.01. From the odds ratios the length of stay 17-341 days carries a 6.752 times the odds of mortality, than a length of stay 6-16 days.

Figure 6.26 Regression Node Output

Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Exp(Est)
Intercept		1	0.1234	0.2284	0.29	0.5890	1.131
Age	21-40	1	-0.5902	0.2989	3.90	0.0483	0.554
Age	41-60	1	-0.2092	0.2891	0.52	0.4692	0.811
Age	61 or older	1	0.1752	0.2921	0.36	0.5487	1.191
Discharge_Department	Internal Medicine	1	-0.1337	0.1955	0.47	0.4939	0.875
Disease_Class	Circulatory	1	-0.4136	0.5356	0.60	0.4400	0.661
Disease_Class	Congenital	1	-1.1430	0.7537	2.30	0.1294	0.319
Disease_Class	Eye and Ear	1	-0.5042	0.5956	0.72	0.3973	0.604
Disease_Class	Gastrointestinal	1	0.3806	0.7446	0.26	0.6092	1.463
Disease_Class	Miscellaneous	1	0.5120	0.3786	1.83	0.1763	1.669
Disease_Class	Muscle	1	0.9644	0.6053	2.54	0.1111	2.623
Disease_Class	Neoplasm	1	0.0536	0.3308	0.03	0.8712	1.055
Disease_Class	Pulmonary	1	-0.4396	0.5382	0.67	0.4140	0.644
Gender	Female	1	-0.2189	0.1677	1.70	0.1916	0.803
Length_of_Stay	1-5 Days	1	-1.0193	0.2272	20.13	<.0001	0.361
Length_of_Stay	LOS 17-341 Days	1	1.4645	0.2716	29.07	<.0001	4.326

Odds Ratio Estimates

Effect		Point Estimate
Age	21-40 vs Under 20	0.297
Age	41-60 vs Under 20	0.435
Age	61 or older vs Under 20	0.638
Discharge_Department	Internal Medicine vs Surgery	0.765
Disease_Class	Circulatory vs Urinary	0.367
Disease_Class	Congenital vs Urinary	0.177
Disease_Class	Eye and Ear vs Urinary	0.335
Disease_Class	Gastrointestinal vs Urinary	0.811
Disease_Class	Miscellaneous vs Urinary	0.925
Disease_Class	Muscle vs Urinary	1.455
Disease_Class	Neoplasm vs Urinary	0.585
Disease_Class	Pulmonary vs Urinary	0.357
Gender	Female vs Male	0.645
Length_of_Stay	1-5 Days vs LOS 6-16 Days	0.563
Length_of_Stay	LOS 17-341 Days vs LOS 6-16 Days	6.752

Model Summary

In summary, the decision tree will take the form of 1 or more inputs and 1 target variable. The inputs may be interval, binary or nominal, the target may also be interval, binary, or nominal. To evaluate the decision tree we can use error, lift, English rules, and misclassification rate for a binary, nominal or categorical target variable.

- Model: Decision Tree node

- Decision Tree: 1+ input and 1 target variable
- Input: Interval, Binary, or Nominal
- Target: Interval, Binary, or Nominal
- Evaluation: Error, Lift, English Rules, Misclassification Rate

Experiential Learning Application: Self-Reported General Health

Executive Summary

One study found that people often evaluate their overall health and wellness based on their lived health rather than their experience of biological health. The self-reported general health (SRGH), contains the levels of very good, good, fair, bad, and very bad. SRGH is one of the most commonly utilized measures of health in population health and clinical health surveys, and utilized to compare populations.Nearly 2,000 scientific studies have been conducted using SRGH or general survey question of how you would rate your health. The question is also used internationally, and included as part of the European Organization of Research and Treatment of Cancer Quality of Life Questionnaire. SRGH has been used as an input variable such as in predicting health outcomes such as mortality, or used as a target variable based on inputs such as diagnosis and symptoms (Bostan et al., 2014).

For the experiential learning application you have been provided a sample of 27,446 records with SRGH as the target variable, and setting type, gender, age group, education, number of health conditions, biological health score and lived health score as the input variables. Help your management team answer the following question.

Question: Which factors may influence SRGH?

Dataset File: 6_EL2_SRGH.xlsx

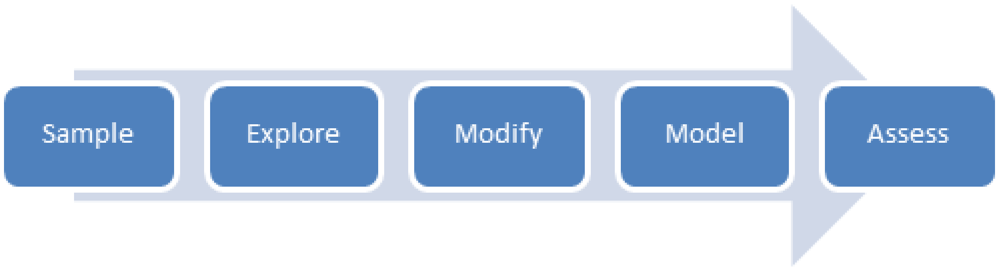
Variables:

- Population Type: 17,739 Community-Dwelling, 9,707 Institutionalized Population
- SRGH: Very Good, Good, Fair, Bad, very bad
- Gender (male, female)
- Age groups (<=65,>65)
- Education (no school, primary-school-incomplete, primary-school-complete, secondary school first step, secondary school finished, professional school medium, professional school superior, University)
- Number of health conditions (0, 1-2, > 2). The health conditions were: Spinal cord injury, Parkinson's, Lateral sclerosis, Multiple sclerosis, Agenesis/Amputation, Laryngectomy, Arthritis, Rheumatoid arthritis or Ankylosing spondylitis, Muscular dystrophy, Spina bifida/hydrocephaly, Myocardial infarction or Ischaemic cardiopathy, Cerebrovascular accidents, Down's Syndrome, Autism and other disorders associated with autism, Cerebral paralysis, Acquired brain damage, Senile Dementia of the Alzheimer Type, Other types of dementia, Schizophrenia, Depression, Bipolar disorder, Pigmentary retinosis, Myopia magna, Senile macular degeneration, Diabetic retinopathy, Glaucoma, Cataract, HIV/AIDS, Rare illnesses, Cancer (only for community dwelling population).
- Biological Health Score 0 (best biological health) to 100 (worst biological health)

- Lived Health Score 0 (best lived health) to 100 (worst lived health)

Follow the SEMMA process for your experiential learning application and provide recommendations. A template has been provided below that can be re-used across future projects.

Figure 6.27 SEMMA Process



Title	**Self-Reported General Health**
Introduction	Provide a summary of the business problem or opportunity and the key objective(s) or goal(s). Create a new SAS Enterprise Miner project. Create a new Diagram.
Sample	Data (sources for exploration and model insights) Identify the variables datatypes, the input and target variable during exploration. Add a FILE IMPORT Provide a results overview following file import: Input / Target Variables Generate a DATA PARTITION
Exploration	Provide a results overview following data exploration Add a STAT EXPLORE Add a GRAPH EXPLORE Add a MULTIPLOT Summary statistics (average, standard deviation, min, max, etc.) Descriptive Statistics Missing Data Outliers
Modify	Provide a results overview following modification Add an IMPUTE
Model	Discovery (prototype and test analytical models) Apply a decision tree model and provide a results overview following modeling. Add a DECISION TREE Model description Analytics steps Decision Tree results (tree model, English rules) Model results (Lift, Error, Misclassification Rate)

	Selection Model
Assess and Reflection	Provide overall recommendations to business Model advantages / disadvantages Performance evaluation Model recommendation Summary analytics recommendations Summary informatics recommendations Summary business recommendations Summary clinical recommendations Deployment (operationalization plan: timeline, resources, scope, phases, project plan) Value (return on investment, healthcare outcomes)

Learning Journal Reflection

Review, Reflect and Retrieve the following key chapter topics only from memory and add them to your learning journal. This may take effort and seem difficult at first, however effortful reflection and retrieval helps builds learning pathways to more easily find the way to and from your existing knowledge in long-term memory. Some difficulties encountered during retrieval help to make the learning stronger and better remembered; effort changes the brain making new connections and pathways, increasing intellectual ability.

For each topic list a one sentence description/definition and example. Connect these ideas to something you may already know from your experience, other coursework, or a current event. This follows our three-phase learning approach of 1) Capture, 2) Communicate, and 3) Connect. After completing, verify your results against your learning journal and ensure all topics are included in your learning journal and update as needed.

Key Ideas – Capture	Key Terms – Communicate	Key Areas - Connect
Payer Anamatics		
Payers		
Claims System and Process		
Claims Forms		
EDI		
Claims Adjudication		
Decision Tree		

Inpatient Mortality Application		
Self-Reported Health Application		

References and Resources

Archer, D., Marmor, T. 2012. Health Affairs. healthaffairs.org/blog/2012/02/15/medicare-and-commercial-health-insurance-the-fundamental-difference/.

Bostan, C. et al. 2014. Biological health or lived health: which predicts self- reported general health better? BMC Public Health.

Chae, Y. M., Kim, Hye S., Tark, Kwan C., Park, Hyun J., Ho, Seung H. 2003. Analysis of healthcare quality indicator using data mining and decision support system. Expert Systems with Applications 24: 167–172.

Change Healthcare. 2017. Inspiring a Better Healthcare System. http://www.changehealthcare.com/about-us/company-overview.

Cheek, K. 2014. An Overview of Analytics in Healthcare Payers. Analytics in Healthcare and Life Sciences.

CMS. 2014. Institutional paper claim form (CMS-1450). https://www.cms.gov/Medicare/Billing/ElectronicBillingEDITrans/15_1450.html.

CMS. 2016. Professional Paper Claim Form (CMS-1500). https://www.cms.gov/Medicare/Billing/ElectronicBillingEDITrans/16_1500.html.

Cognizant. 2014. Cognizant to Acquire TriZetto, Creating a Fully-Integrated Healthcare Technology and Operations Leader. http://investors.cognizant.com/2014-09-15-Cognizant-to-Acquire-TriZetto-Creating-a-Fully-Integrated-Healthcare-Technology-and-Operations-Leader.

Epic. 2017. Managed Care. http://www.epic.com/software#ManagedCare.

Heilbronn, E. 2014. US News. http://health.usnews.com/health-news/health-insurance/articles/2013/12/16/top-health-insurance-companies.

Hirsch, L. and Sherman, A. 2018. Walgreens and AmerisourceBergen deal talks have cooled as takeover looks unlikely. CNBC. https://www.cnbc.com/2018/02/27/walgreens-and-amerisourcebergen-deal-talks-of-cooled-.html.

KFF. 2016. Employer Health Benefits Survey. Kaiser Family Foundation. http://www.kff.org/report-section/ehbs-2016-section-ten-plan-funding/.

Klimberg, R., McCullough, B.D. 2013. Fundamentals of Predictive Analytics with JMP. SAS Institute, Cary NC.

Livingston, S. 2016. Modern Healthcare. http://www.modernhealthcare.com/article/20161012/NEWS/161019970.

MedicalBillingandCoding. 2017. How Does Commercial Health Insurance Work? http://www.medicalbillingandcodingu.org/how-does-commercial-health-insurance-work/.

McDowell, W., et al. 2018. Cigna to Acquire Express Scripts for $67 billion. https://www.cigna.com/newsroom/news-releases/2018/cigna-to-acquire-express-scripts-for-67-billion.

McKesson. 2017. Claims Management Technology. http://www.mckesson.com/health-plans/network-and-financial-management/clinical-claims-management/claimsxten/.

Moeller, M. 2009. Manage Medical Advances With Automated Prior Authorization. Managed Healthcare Executive.

Murphy, T. 2017. Anthem gives up Cigna bid, vows to fight on over damages. Associated Press. https://finance.yahoo.com/news/anthem-gives-cigna-bid-vows-181247826.html.

NCSL. 2017. Health Insurance: Premiums and Increases. National Conference of State Legislatures. http://www.ncsl.org/research/health/health-insurance-premiums.aspx.

PHDSC. 2007. Public Health Data Standards Consortium. Users Guide for Source of Payment Typology. http://www.phdsc.org/about/committees/pdfs/SourceofPaymentTypologyUsersGuideOct2007.pdf.

Pearson, B. 2018. 5 Ways A Walmart-Humana Partnership Could Change Everyone's In-Store Experience. Forbes. https://www.forbes.com/sites/bryanpearson/2018/04/12/5-ways-a-walmart-humana-partnership-could-change-everyones-in-store-experience/#2cc84a7a58a7

Porterfield, A., Engelbert, K., & Coustasse, A. 2014. Electronic Prescribing: Improving the Efficiency and Accuracy of Prescribing in the Ambulatory Care Setting. Perspectives in Health Information Management, 11(Spring), 1g.

Ramsey, L. 2018. CVS and Aetna's megamerger could get blocked, and we'll soon get a hint of what will happen next. Business Insider. http://www.businessinsider.com/cvs-health-aetna-merger-potential-antitrust-challenges-to-the-deal-2018-4

Tricare. 2017. About Us. https://www.tricare.mil/About.

SIIA. 2015. Self-Insured Group Health Plans. Self-Insurance Institute of America. https://www.siia.org/i4a/pages/index.cfm?pageID=4546.

Trizetto. 2017. Claims Processing. http://www.trizetto.com/Provider-Solutions/Health-Systems/Claims-Processing/?utm_source=facebook&utm_medium=social&date=091514.

Tracer, Z., McLaughlin, D., Harris, A.M. 2017. After Two Megadeals Blocked, Health Insurers Plot Next Moves. Bloomberg. https://www.bloomberg.com/news/articles/2017-02-09/anthem-s-bid-for-cigna-blocked-by-judge-as-anticompetitive.

TM Floyd & Company. 2006. Health Claims Processing in the US. TM Floyd & Company.

VFW. 2017. What is CHAMPVA and who is eligible? http://faq.vfw.org/faq/index.php?View=entry&EntryID=445.

VA.gov. 2017. Veterans Health Administration. https://www.va.gov/health/.

Woodside, J.M. 2007. EDI and ERP: A Real-Time Framework for HealthCare Data Exchange. Journal of Medical Systems.

Woodside, J.M. 2010. A BI 2.0 Application Architecture for Healthcare Data Mining Services in the Cloud. The World Congress in Computer Science, Computer Engineering & Applied Computing - International Data Mining Conference.

Yu, T., Hou, Y., Lin, K., Chung, K. 2014. Is it possible to identify cases of coronary artery bypass graft postoperative surgical site infection accurately from claims data? A multi-model comparison study over 2005–2008. BMC Medical Informatics and Decision Making, 14:42