MS ADS Portfolio

Data Warehouse
- Introduction to Data Warehousing
- Data Warehouse Architectures
- Building the Data Warehouse
- Dimensional Modeling
- Dimensional Model Development
- Introduction to ETL
- ETL Development
- Managing the Data Warehouse
- Business Intelligence
- Big Data and the Data Warehouse

Tools and techniques used: Data profiling, High level Dimensional Modeling, Detailed dimensional modeling, ETL (Extract, Transform,Load), Business Inteligence
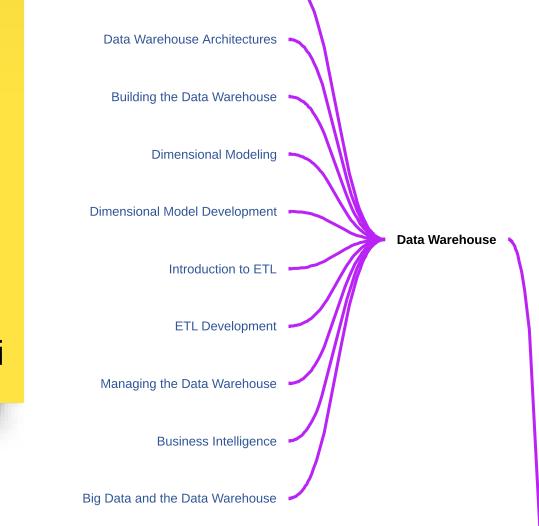*Obongo Omondi*

Final Project: Created a datawahouse that facilitated corporate analysis across businesses and conducted a BI analysis of those businesses.
Obongo Omondi

Lessons Learned
- Expert work with SQL
- Understanding Data warehousing architectures
- Building a modern data warehouse
- Dimensional Modeling and development
- ETL Development
- Business Inteligence
- Big Data and the Data Warehouse
Obongo Omondi

(DATA WAREHOUSE)

Quantitative Reasoning in Data Science
- Essential Statistical Concepts and Vocabulary
- Events and Probabilities
- Sampling and Long-Run Probabilities
- The Logic of Inference With Confidence Intervals
- Bayesian and Traditional Hypothesis Testing
- Comparing Groups and Analyzing Experiments
- Associations Between Variables
- | Linear Multiple Regression
- Logistic Regression
- Analyzing Change Over Time

Tools and techniques used: Loading and Unloading Data into R-studio, Descriptive statistics and distributions, Sampling techniques, sampling distributions, statistical inference, Time Series/Trend analysis, Bayes Factor Analysis
Obongo Omondi
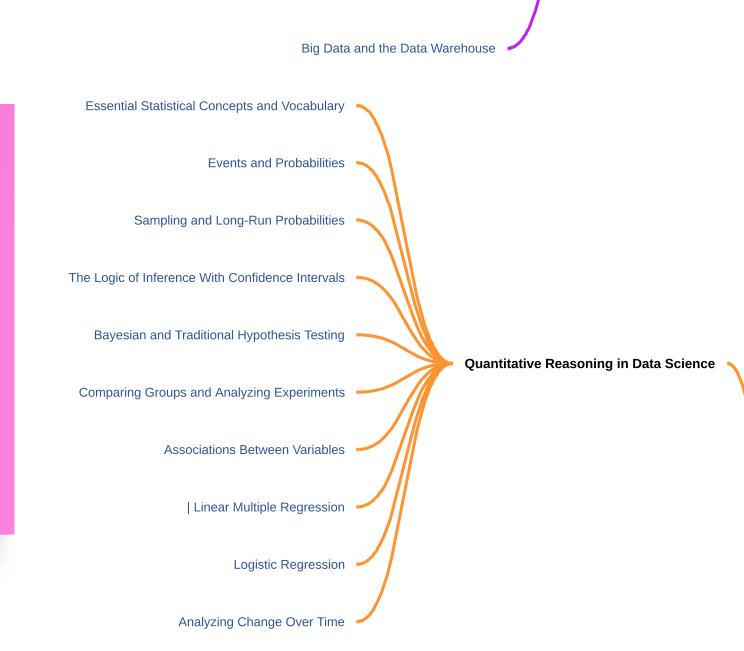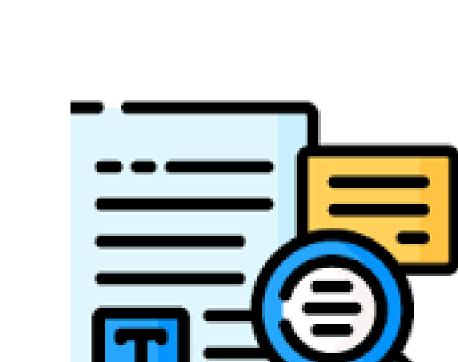
Final Project: I conducted a statistical analysis of vaccine reporting data from kindergartens in California - analyzing proportions of private and public participation. Investigated conditional, medical and religious factors in predicting whether a school is private or public. Investigated conditional percentages based on percentage of specific vaccines that are missings. Used both Frequentist and Bayesian approaches to all predictionss.
Obongo Omondi

Lessons Learned
- Preparing data for analysis, including screening data, dealing with missing data, doing data transformations
- Testing assumptions that data must meet for analyses and inferences to be reasonable
- Interpreting data analysis results and outputs and communicating them to otherssusing language that acurtely describs uncertainty
- Leaving documentation / provenance trail for other analysts to follow and reproduce your work
Obongo Omondi

(Quantitative reasoning)

Text Mining
- Introduction to Text Mining
- Document Vectorization
- Corpus Analysis
- Naive Bayes for Text Categorization
- Text Classifier Evaluation and Human Annotation
- Scikit-learn
- SVMs for Text Categorization
- Document Clustering and Topic Modeling
- Literature Review on Text Mining Applications/Project Idea Presentation
- New Topics in Text Mining/Project Clinic and Presentation

Tools and techniques used: Machine Learning, Natural Language Processing, Statistics, Business Inteligence, information orgazation and access, social behavior analysis, digital humanities.
Obongo Omondi

Final Project: I analyzed public tweets on 2019 democratic presidential debates immediately after the debates. I leverage Twitter posts with public sentiment on policy issues, opitions, support and criticism of the positions taken by the candidates.
Obongo Omondi

Lessons Learned
- Computational methods to find patters in large text corpora,
- How to use Machine learning and Natural Language processing to process text corpora
- How to use statistical design in analysing text tokens
Obongo Omondi

Advanced Database Management
- Course Introduction: Setting the Stage
- Relational Databases: Programming
- Relational Databases: Transactions, Concurrency, and Triggers
- Relational Databases: Performance, Security, and noSQL
- Hadoop: HDFS, MapReduce and YARN
- Hadoop: Pig, Hive and Hcatalog
- Hadoop: Hbase, Impala, and Other Tools
- NoSQL: MongoDB and Redis
- NoSQL: Cassandra
- NoSQL: Kafka and KSQL

Tools and techniques used: Kafka, Cassandra, Hadoop, Mongo, SQL, CAP Theorem, .
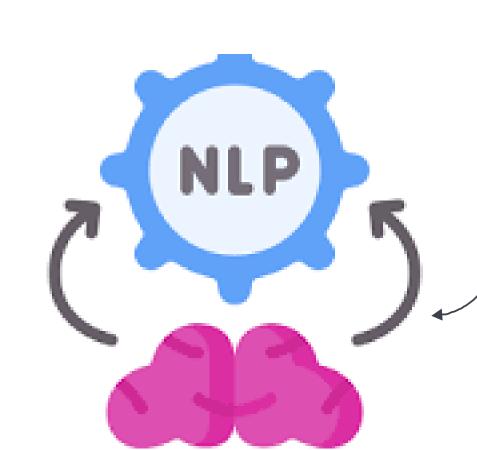*Obongo Omondi*

Homework Series: Hands-on work writing SQL and NoSQL variants for Streaming (Kafka), document (Mongo), columnar (Cassandra), relational (SQL) database systems
Obongo Omondi
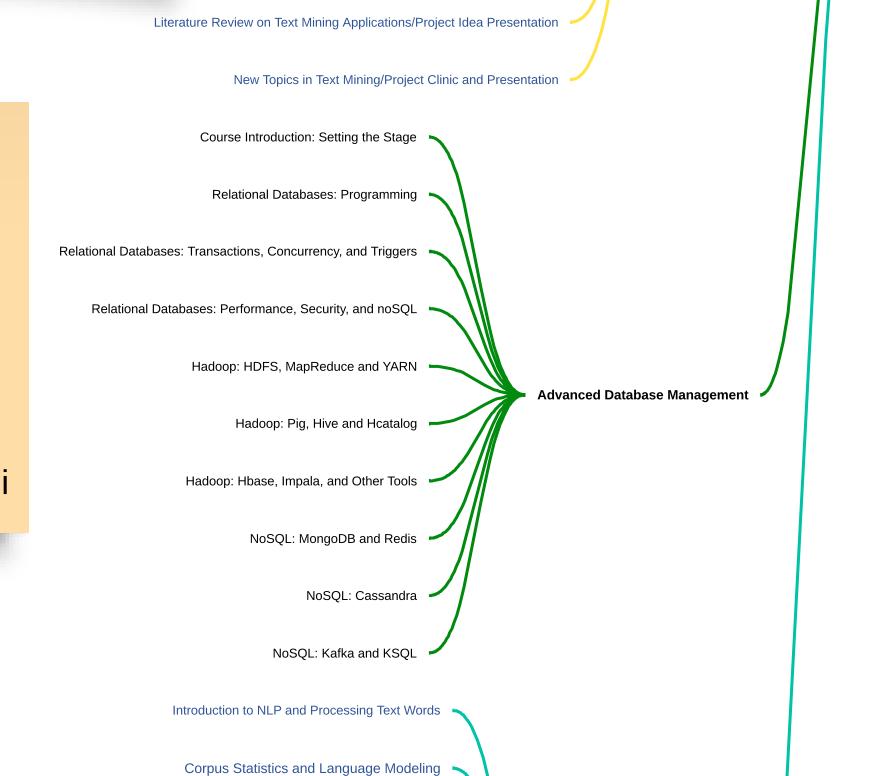
Lessons Learned
- How to work with relational, document, key-value, columnar, and streaming database systems through the lense of CAP theorem.
- Working with relational, Hadoop and NoSQL databases.
Obongo Omondi

Natural Language Processing
- Introduction to NLP and Processing Text Words
- Corpus Statistics and Language Modeling
- Regular Expressions, Morphology, and Processing Text Files
- POS Tagging and Introduction to Machine Learning
- Context-Free Grammars (CFG) and Parsing
- Semantics
- Discourse and Dialogue
- Sentiment Analysis
- NLP Applications: Information Extraction, Machine Translation, and Summarization
- NLP Applications: Information Retrieval, Question Answering, and Conversational Agents

Tools and techniques used: NLTK, ScikitLearn, Classification, Model training, feature engineering, visualization, Sentiment/Subjectivity Analysis, Tokenization, Lemmatization, Stemming, Regular Expression, POS tagging
Obongo Omondi

Final Project: I conducted an analysis of Kaggle Movie Review Data based on reviews from rotten tomatoes website. Deployed a training data consisting of 156K phrases for analysis.
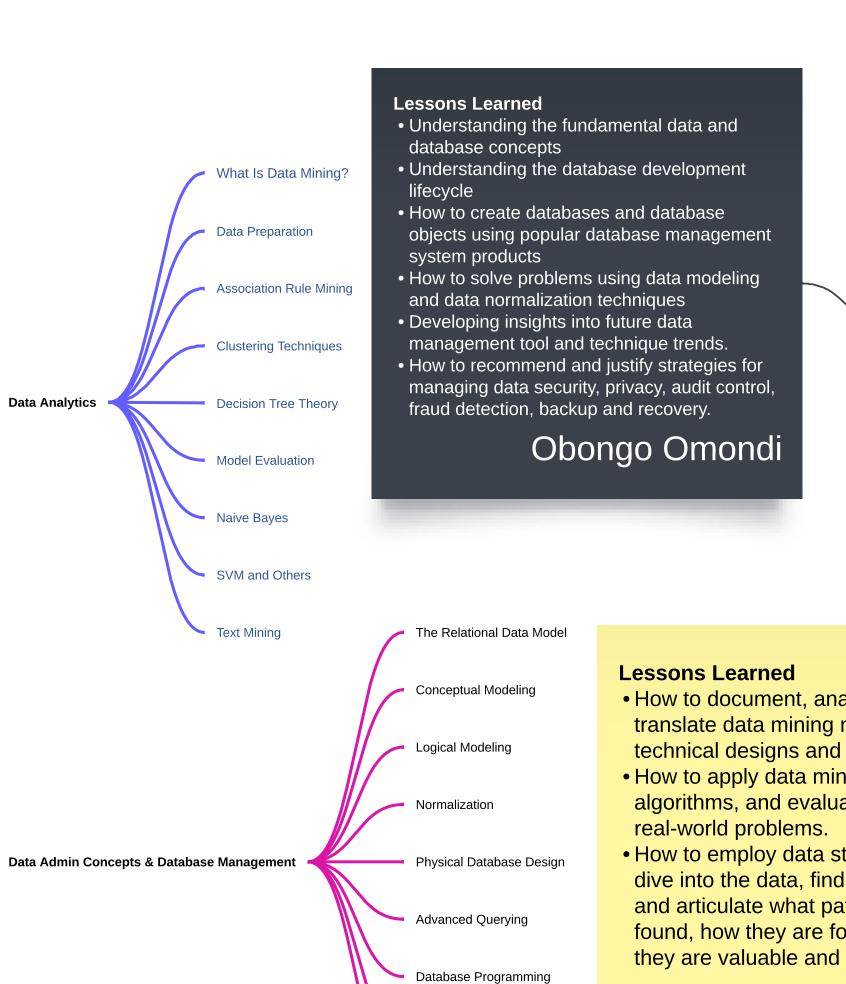*Obongo Omondi*

Lessons Learned
- How to use the NLTK library for semantic and linguistic computational interpretation of text
- Understanding and working with text corpus
- Preprocessing of large text corpus
- Tokenization using different approaches such as the Sklearn Count vectorizer, NLTK Word Tokenizer, Wordpunct Tokenizer,
- Using the Regular Expression library of Python
- Stemming and Lemmatization
- Feature selection
- Sentiment/Subjectivity analysis
- Parts of Speech Tagging
- Classification with Naive Bayes
Obongo Omondi

(NLP)

Data Analytics
- What Is Data Mining?
- Data Preparation
- Association Rule Mining
- Clustering Techniques
- Decision Tree Theory
- Model Evaluation
- Naive Bayes
- SVM and Others
- Text Mining

Lessons Learned
- Understanding the fundamental data and database concepts
- Understanding the database development lifecycle
- How to create databases and database objects using popular database management system products
- How to solve problems using data modeling and data normalization techniques
- Developing insights into future data management tool and technique trends.
- How to recommend and justify strategies for managing data security, privacy, audit control, fraud detection, backup and recovery.
Obongo Omondi

Final Project: I worked on a Pre-diabetes prediction models with machine learning using Kaggle data taken from hospitals in Frankfurt, Germany. Conducted training and splitting of the dataset, and employed different machine learning algorithms to model and anlyze the data such as Naive Bayes, Decision Trees, SVM polynomial, SVM Linear, SVM radial Kernel, and Random Forest. I provided discussion on accuracy of each approach.
Obongo Omondi

Tools and techniques used: SQL-server, Visio, draw.io, Access, data modeling, SQL, data modeling, relational data models, normalization, ERD, Transaction Management, Concurrency, distributed databases, mulitier client/server architectures, web based database applications, data warehousing, NoSQL.
Obongo Omondi

Data Admin Concepts & Database Management
- The Relational Data Model
- Conceptual Modeling
- Logical Modeling
- Normalization
- Physical Database Design
- Advanced Querying
- Database Programming
- Database Administration
- User Interface Design

Lessons Learned
- How to document, analyze, and translate data mining needs into technical designs and solutions.
- How to apply data mining concepts, algorithms, and evaluation methods to real-world problems.
- How to employ data story telling and dive into the data, find useful patterns, and articulate what patterns have been found, how they are found, and why they are valuable and trustworthy.
Obongo Omondi

Final Project: I worked on a Pre-diabetes prediction models with machine learning using Kaggle data taken from hospitals in Frankfurt, Germany. Conducted training and splitting of the dataset, and employed different machine learning algorithms to model and anlyze the data such as Naive Bayes, Decision Trees, SVM polynomial, SVM Linear, SVM radial Kernel, and Random Forest. I provided discussion on accuracy of each approach.
Obongo Omondi

Tools and techniques used: Data mining, data preparation, concept description, association rule mining, classification, clustering, evaluation and analysis. Waka, Rapid Miner, SVM, Random Forest, Decision Trees, Naive Bayes.
Obongo Omondi

(DATABASE ADMINISTRATOR)

Applied Data Science
- Introduction
- Rows and Columns
- Writing Functions and Descriptive Stats
- Sampling
- Accessing Data
- Pictures Versus Numbers
- Map Mashup
- Linear Modeling
- Data Mining
- Text Mining

Lessons Learned
- Basic to intermediate use of the R-Programming languge by writing simple and complex functions. Focus on descriptive stats
- How to access data with R and statistical Sampling techniques used in industry
- Linear modeling with R
- Data mining and text mining with R
- Data visualiztion with R packages
Obongo Omondi

Homework/Assignments: A series of incremental learning how to use R for statistical analysis via homework assigments of incremental complexity.
Obongo Omondi

Tools and techniques used: Extensive use of the R programming Language for statistical analysis.
Obongo Omondi

(R)

Scripting for Data Analysis
- Data Pipeline and Python Language Basics
- Arrays, Functions, and Categorical Summarization
- Stacking and Unstacking Data
- Semi-structured Data
- Mongo Database, JSON From RSS
- Processing Twitter and Facebook
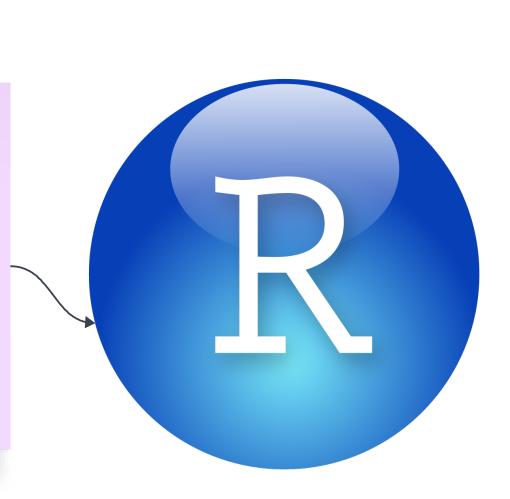- Unstructured Data
- Network Structures
- Booleans and Dictionaries
- Exploring and Transforming Data for Structured Data

Lessons Learned
- How to work with Structured, Semi-structured and unstructured data
- How to handle and process structured numeric text data from spreadsheets and databases
- How to use data obtained through standard data exchange formats such as HTML or XML from web pages or JSON from web-based APIs
- How to use data obtained by pattern matching from text or log files.
Obongo Omondi

Final Project: Analysis of Tweeter Data: I analyzed tweets based on the "#KarenStrikesAgain" hashtag. I logged into the twitter API and searched for for tweets based on the "karen" meme as a search term. I collected raw tweets and placed them into a neat pandas dataframe, then summarized the top 20 tweets based on location and date of tweets. I further performed analysis of the tweets over time, breaking down the tweets by years, months, weeks and hours. I answered questions such as when and where the tweets surged or waned and provided visualizations using pandas plotly to demonstrate this. Also provided sentiment analysis of the tweets which turned out to be mostly negative - only 5% of the KarenStrikesAgain had a positive connotation.
Obongo Omondi

Tools and techniques used: Python programming, pandas, plotly, matplotlib, seabon,sentiment analysis, text mining with NLTK, parts of speech tagging, working with csv, data prep before analysis.
Obongo Omondi