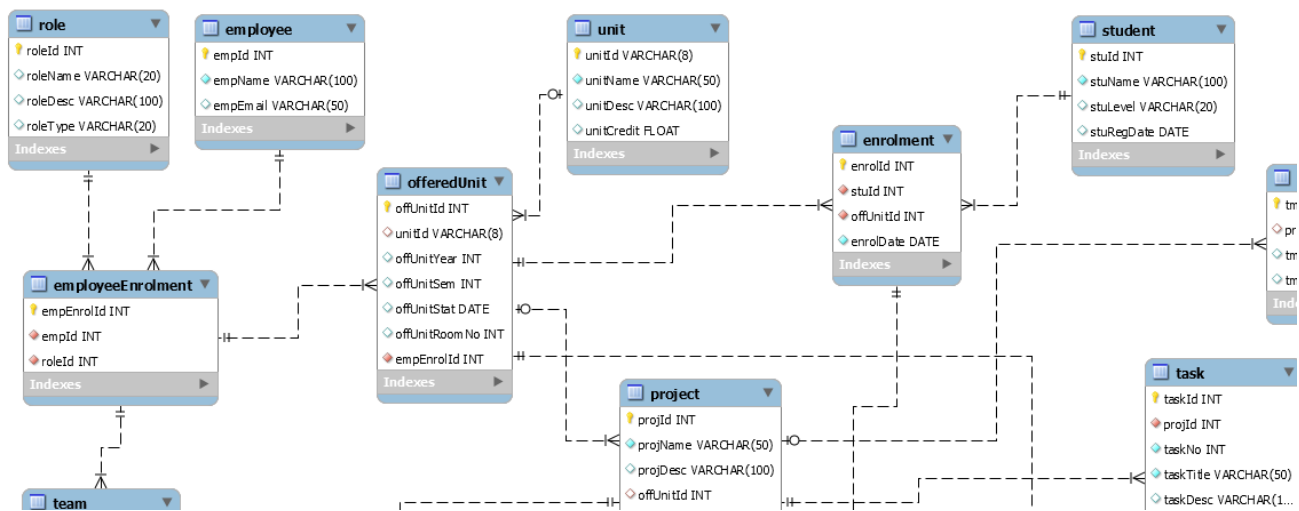# PORTFOLIO

Kianoosh Keshavarzian

Data Scientist

Completed Projects Samples, Recent Qualifications and Certificates

Completed Projects Samples

Database Design and Implementation

Client:

Swinburne University of Technology

Description:

Successfully delivered an advanced, three tier database project for tutorial segment of Swinburne University of Technology based on MySql database and ASP.NET (VB) frontend which was published on AZURE. The project was performed by a group of five and was chosen as the best one out of the three.

Technologies used:

MySql, ASP.NET (VB), AZURE, HTML, CSS, JavaScript
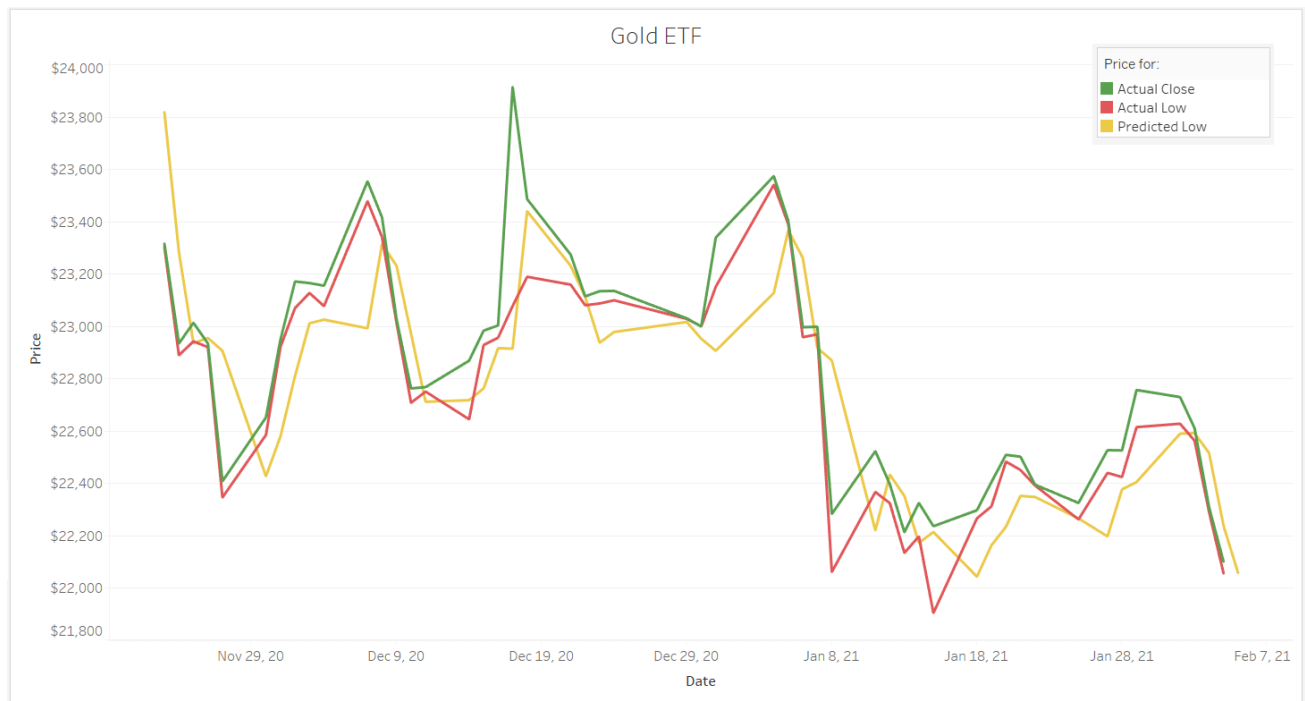
Teamwork:

Group of five

Documents delivered:

Entity Relationship Diagram (ERD), Systems Requirement Specification (SRS), Use Cases, Test Cases, User Documentation, Business Rules & Traceability Matrix

Duration:

Six Month

Stock Market Price Prediction using LSTM

Client:

Bonafide Advice

Description:

Developed Python code to use API to receive businesses statistics and fundamental analysis from Alpha Vantage website, predict the future of the market by implementing a hybrid LSTM model to include variety of markets and handover better predictions on the price movement. Then, populate the Oracle database with analytic results and make webpages connected to the database to deliver the results to the customers.

Technologies used:

API, Python, tensorflow, sklearn, pandasql, Oracle

Teamwork:

Individual

Documents delivered:

User Manual, Test Cases

Duration:

Four Month

Costumer Reviews Prediction using Deep Learning

Client:

Upwork Client

Description:

Developed an NLP Deep Learning model in Python to predict the number of like a review will receive in a restaurant website. The model is using the keras classifier and was able to achieve accuracy of 73% and precision of 69% for over 1000 reviews. Once data preparation function was ready, other classification models such as decision tree, kernel SVM, KNN, logistic regression, naive bays and random forest was suggested to the client to be used to check the reliability of the results. A couple of these models was added to the code as a bonus.

Technologies used:

Python, Keras classifier, nltk, re, corpus, Porter Stemmer
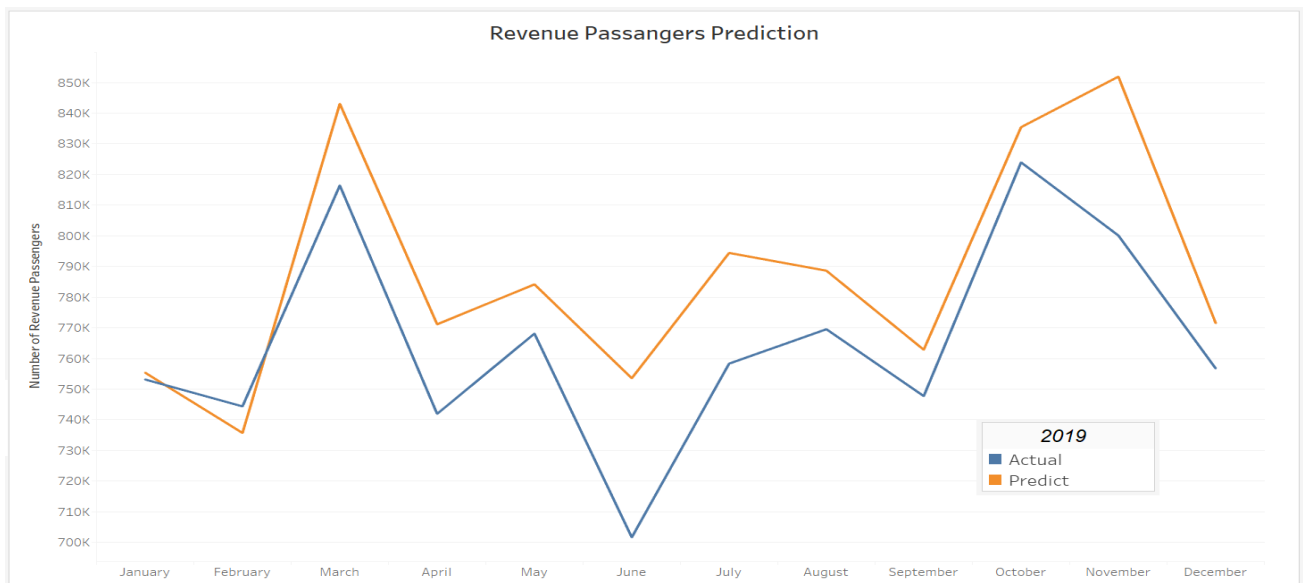
Teamwork:

Individual

Documents delivered:

ReadMe file, Model description

Duration:

Seven Weeks

Number of Revenue Passenger Prediction using LSTM

Client:

Jetstar Airways

Description:

Predicting the number of revenue passengers for the upcoming month using Multivariate LSTM. The variety of the training set was very limited as the dataset was only contained of the revenue passengers' number for all the cities involved. The number of missing data was also considerably high.

Technologies used:

Python, Tensorflow-Keras, sklearn, pandasql, Tableau

Teamwork:

Group of two

Documents delivered:

Model Description, Tableau Dashboard

Duration:

Two Month

Candlestick Chart conversion to Dataset

Client:

Upwork, SEST

Description:

Developed a package to convert stock market candlestick chart into dataset. The code is using Optical Caracter Recognition (OCR), PIL and Pytesseract packages to read the axes values from the image. It also applies a manual calibration process to remove the background and unnecessary parts of the image. It is able to retrieve data from the charts with only 0.02% deviation and slash a significant portion of spending on raw data.

Technologies used:

API, Python OCR, PIL, Pytesseract, Data Visualization, Image Processing

Teamwork:

Individual

Documents delivered:

Stock Market Brokers Guide, Image Preparation Guide

Duration:

Three Month

Airbnb Investment Analysis

Radar chart comparing Cluster 1, Cluster 2, and Cluster 3 across: Entire Apartment, High Availability, Multi-listing Host, Top20 Multi-listing Host, Estimated Income, Occupancy, Airbnb Density.

Client:

Private Investor

Description:

Attained impressive accuracy in two unsupervised clustering models to categorize the suburbs in Victoria based on their profitability for investing on Airbnb market whereas the dataset has been extremely dirty. Excel functions to cleanse the dataset and Python's KMeans and Hierarchy clustering models to cluster the suburbs were used.

Technologies used:

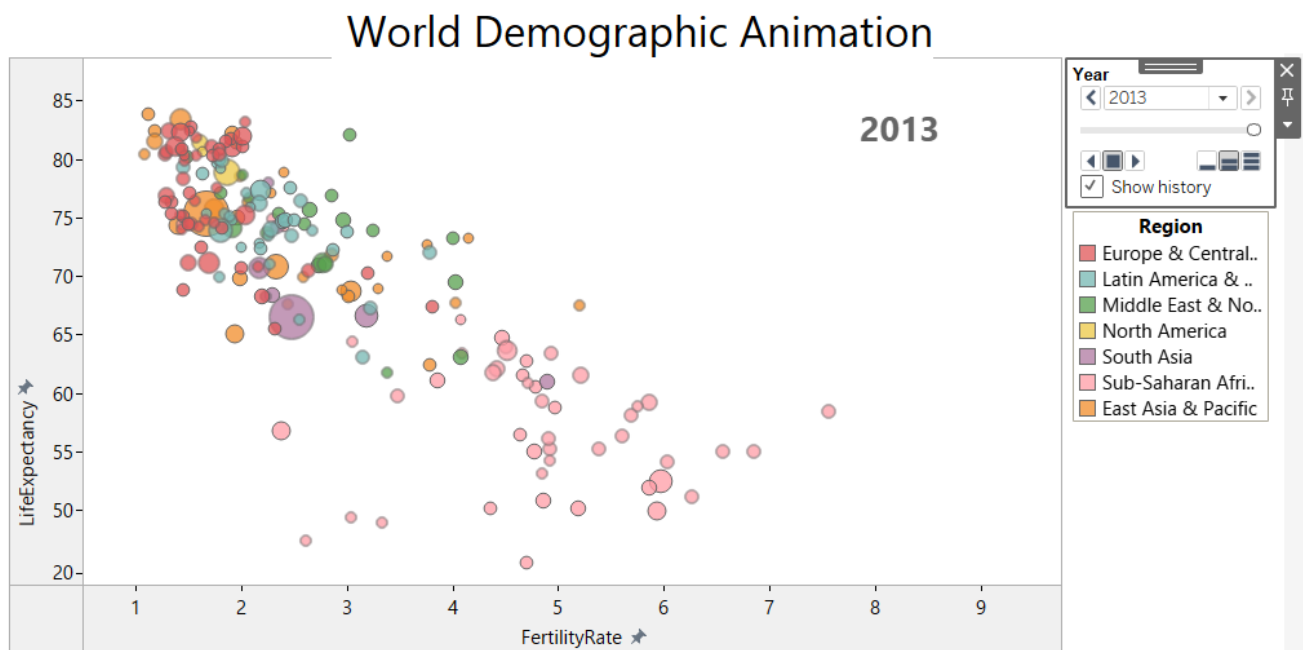Excel, Python, Scikit Learn, KMeans, Scipy

Teamwork:

Group of three

Documents delivered:

Research Report

Duration:

Three Month

Live Data Visualization in Tableau

Client:

Upwork Client

Description:

An advanced dynamic PowerPivot as well as a live animated Tableau visualization on a dataset containing population, fertility rate, life expectancy, country, region and year was performed. The animation moves along the year attribute and visualize the population via size and the region via color as well.

Technologies used:

Excel, PowerPivot, Tableau
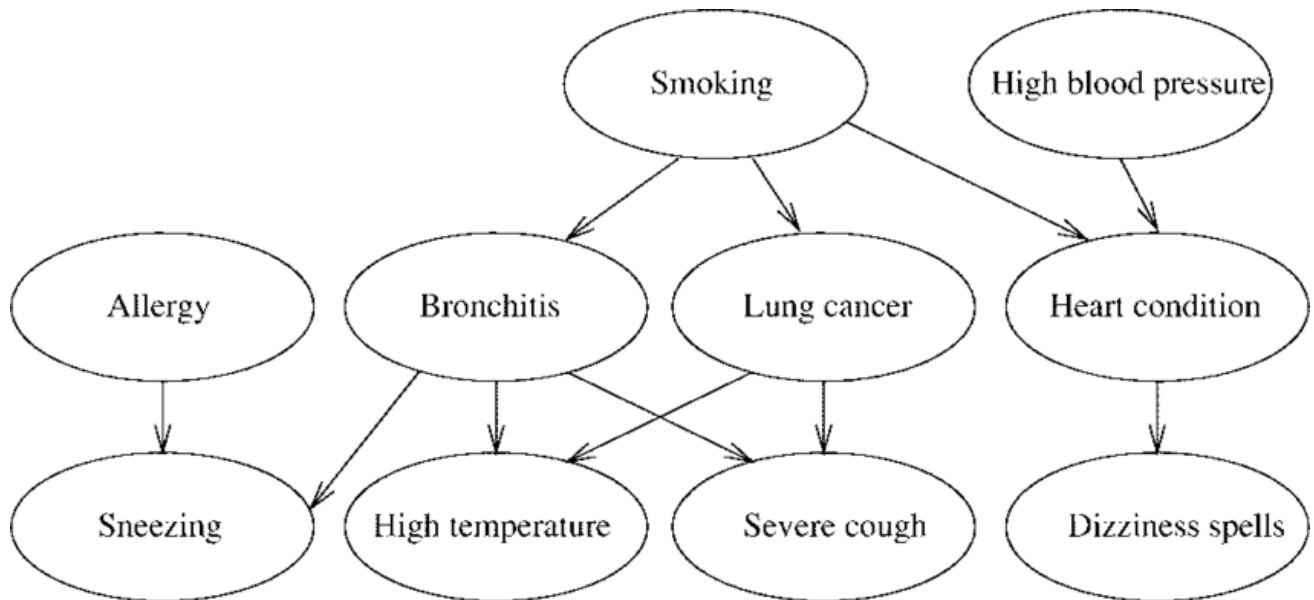
Teamwork:

Individual

Documents delivered:

ReadMe file, Animated Video

Duration:

One Month

Association Rule Learning in Medicine using Apriori and Eclat

Client:

Private General Practitioner

Description:

Successfully designed a model to find numerical meaningful relations between symptoms stated by the patients. The model uses apriori function of apyory package in Python or eclat function of arules library in R and output the categories of the symptoms found as the rules. The model can also be used on any dataset containing different dimensions of one quantity such as items in one supermarket transaction.

Technologies used:

Python, R, apyori, arules

Teamwork:

Individual

Documents delivered:

Model Description

Duration:

One Month