SUM YU NG

+ 44 07400 751979
+ 852 5426 4279
ruby520sab@gmail.com
https://sumyu199.github.io/sumyuprofile/index.html
https://www.linkedin.com/in/sumyu-ng-b64a321b0/
https://github.com/sumyu199

Education

University of Nottingham (UK) 2018 - 2021
Bachelor of Financial Mathematics
Risk, Information & Insurance (70)
Discuss behavioural approach and traditional approach to risk so as to evaluate the role of insurance
International Finance (73)
Explore the foreign exchange market and discuss management of Exposure of foreign exchange

University of Nottingham (UK) 2017 - 2018
Engineering and Physical Sciences Foundation Programme
Average score:87

Tsuen Wan Public Ho Chuen Yiu Memorial College (HK) 2011 – 2017
Hong Kong Diploma of Secondary Education
Maths 5| Chinese 4 | Liberal Studies 4

Skills

Tools & Technologies

- Analyse Data with Python (with project)
- Build a Machine Learning Model with Python (with project)
- Analyse Data with SQL
- Microsoft Excel
  Linkedin Excel 2016 Essential Training with NASBA
  Pass the exam with at least 70%
- Google Analytics
  Google Analytics Individual Qualification
- HTML & CSS (Personal Website)
- Rstudio (With Project)
- Tableau (With Project)
  Google Course: Share Data Through the Art of Visualization (Grade Archieved: 85)

Project

Following projects' details are all shown on my personal website

1. Data Preparation and Visualisation (Jupyter Notebook)
   This project is a data preparation task of Quantium Data Analytics Virtual Experience Program and the database files included the customer purchase data and transaction data.
   - To start with, **data cleaning** and **data merging** are required.
   - To explore the data, I used **malplotlib** to visualize the data and defined the target customer segments so as to **reduce the dimension**.
   - Finally, I **deep dived into those segement** to investigate the preference of our target customer so as to develop an efficient market strategy.

   Inspired by this task, I decided to create a **Tableau dasbroad** to tell the story of the data with different charts.

2. Techincal Analysis (Python)
   This application is using YahooFinance as the data resource and **streamlit** as the framework of the application to process the technical analysis of stock in a shareable **application**.
   - To collect the data that the user interested only, I created a datareader function with **pandas** datareader to take the input parameters from users and retrieve data from YahooFinacne.

- Once the data is collected, the close price of the ticker will be analysed with **Bollinger Bands with RSI, Moving Average Convergence Divergence and On-balance Volume**.In these functions, **window functions** are frequently used. Also, all the results are plotted by **plotly**.
- Finally, the next day close price is predicted by **Support Vector Regression** from sklearn with Radial Basis function kernel, Polynomial regression and Linear Regression.

3. Logistic Regression Predictor (R studio)
This project is to build a predictor model for the **classification** of received a medal or not and the data is included athlete information of Olympics 2012
- To ensure there is no NA value that will affect the future calculation, I filled all the NA values with mean and transform the Medal column (label) into binary structure.
- For the features, I visulaised the data and applied **t-test** on both non-medalists and medalists to prove the hypothesis that medalists have bneneficial features to win
- Next, the data is **split into training and testing sets**.
- Finally, I decided to use logistic regression glm to build the model as **logistic regression** is suitable for classification. The features are decided by applying **z-test and p-value** and the classification prediction **accuracy is about 81%.**

4.Decision Trees Predictor (Jupyter Notebook)
This project is to use the historical data of titanic passengers to create a predictor to predict whether the passenger survived or not.
- To start with, I used **lambda and numpy** to clean the data and examinate the **correlation** of features.
- Next, I selected different sets of features to create several models and split the data into **reproducible training and testing sets**.
- Finally, I created both **random forest classifier and decision tree classifier** and as the decision tree with the best depth has a higher score than forest, so decision tree is decided to be the final model. This model is **scored 0.7799** in Kaggle.

Relevant Experience

Temp. Assistant |Data Entry and Activity Organisation June 2018 – Sept 2018

United Social Services Centre (HK)
- Orgainsed data of service users in Excel
- Contacted service users and replied their query
- Organised activities with my team

Other jobs have also included:
Event helper of exhibition, part-time tutor of primary school children

These work experience help me to develop Excel skill, teamwork skill, customer service skill and communication skill.

Languages

- Cantonese
- Mandarin
- English
  Coursera GTC Statement of Accomplishment
  Contributed at least 15,000 words of high quality transaltation in Coursera.