Shiyu Shi

College Park, MD,20740
▪ (240) 615-6685 ▪ shiyu.shi@marylandsmith.umd.edu ▪ LinkedIn ▪ Github ▪ Website

PROFILE

Specialties: Machine Learning, Statistical Exploratory, Data Visualization, Predictive Modeling, A/B Testing
Skillsets: SQL, Python (pandas, numpy, sklearn, Matplotlib, seaborn), R (dyplr, ML packages, ggplot), Tableau, Excel
Big Data Tools: PySpark, Hadoop MapReduce, AWS (Amazon Web Services)
Certification: AWS Solution Architect - Associate (08/2020)

PROFESSIONAL EXPERIENCE

Pilot Freight Service | Logistic Analytics | Snowflake, SQL, Python, Excel, Tableau **College Park, MD**
Data Analyst, Capstone Project Sep. 2020 – Dec. 2020

- Data Manipulation: Extracted and cleaned 14 billion+ data from linehaul related databases using SQL and Python.
- Mid-Week Strategy: Designed mid-week linehaul shipping strategy and conducted simulation on logistic network. The results show a 42% increase in network volume.
- Profitability Analysis: Developed a cost and revenue model with new midweek operation in Excel to analyze real time variance in volumes and truckload expenses, which can save $250,000 annually.
- Data Visualization: Designed and Implemented interactive dashboards that highlight trends and unprofitable stations using Tableau, SQL and Excel, presenting a list of 20 stations for further actions.

NIKE Sports China | HR Business Analytics | Excel, SAP **Shanghai, CHN**
HRBP Analyst Intern, Global Operation & Technology Team Dec. 2018 – May. 2019

- Maintained people database and monitored ongoing talent development initiatives for five business departments.
- Provided operational analysis of BP functions by analyzing HR key metrics and visualizing team-based results.
- Reduced 80% time consumed by monthly routine work by building Excel models.

ANALYTICS PROJECT

Deloitte NLP Analysis over Tweets about COVID-19 | Python, NLP, Clustering Jul. 2020

- Implemented Word Frequency, Topic Clustering and Sentiment Analysis model based on tweets of United States citizens from both nation-wide and state-wide perspective.
- Visualized results leveraging WordCloud and plotly to compare sentiment bias across United States.

Analysis and Prediction of Movies | Python, NLP, Predictive Modeling Mar. 2020

- Incorporated two Kaggle datasets of 5,000+ movies and wrangled integrated data for machine learning.
- Executed words of interest analysis in time series by drawing word clouds to obtain representative words in movie overview for different level of gross and visualizing frequency of WOI in years.
- Constructed a demo that displays expected gross class by comparing test accuracy of nine classification models.

Prediction of Booking Rate for Airbnb | R, Classification, Machine Learning March. 2020

- Extracted key words of description with LDA (Latent Dirichlet Allocation) to create meaningful factor features for prediction.
- Completed data munging and feature engineering on a dataset of 100,000+ observations.
- Trained nine classification models and applied the best on unlabeled dataset, achieving an accuracy of 83%.

Restaurant Recommendation Platform Design | SQL, Tableau, R Sep. 2019

- Built a relational database containing 1,000+ restaurant in College Park through Microsoft SQL Server and generated dynamic visualizations using Tableau.
- Created a live web interface and word clouds with integrated database and visualizations to facilitate quick and matching restaurant searching by using R Shiny.

EDUCATION

University of Maryland, College Park **College Park, MD, USA**
Master of science, Business Analytics, GPA: (4.0/4.0) Aug. 2019 – Dec. 2020
Award: Terrapin Scholar

Shanghai University of Finance and Economics **Shanghai, CHN**
Bachelor of Management, International Business Sep. 2015 - Jun. 2019
Award: Third-Class People's Scholarship in China (Top 10%)