Aakash Sheth

Phone: 813-539-9401 |**E-mail**: aakashsheth65@gmail.com | GitHub | LinkedIn

PROFESSIONAL SUMMARY

Data Scientist and a certified **Six Sigma Green Belt** professional having a master's degree in **Industrial Engineering** with **2+ years** of hands-on experience leveraging **statistics, data mining** and **machine learning models** to uncover **trends**, **develop forecasting models** and create **dashboards** & **visualizations** to support executive decisioning. Ability to initiate and drive end to end project to completion with minimal guidance. **Expert** in **Excel, VBA, Python, SQL** technologies.

EDUCATION

University of South Florida, Tampa, FL CGPA: 3.76/4.0
Master of Science in Industrial Engineering May 2019

SKILLS

- Data Science: **Python with (NumPy, Pandas and Matplotlib, Django, Pyspark, Scikit learn), Tableau, Git, GitHub, Six Sigma.**
- Programming Technologies: **Python, SQL, VBA, HTML/CSS.**
- Machine Learning: **Regression, Decision Tree, Random Forest, Neural Networks, Ensembles, Gradient Boosting, Regularization, KNN, Tensorflow, Monte Carlo Simulations, ANOVA**

WORK EXPERIENCE

Analyst Data Science, RedPerit LLC (Tampa, FL) June 2018- Present
Tools: Python, Tableau, Excel (VBA, Pivot Table, VLOOKUP), SQL, CRM.

- Creating python functions to carry out **web scrapping**, **data manipulation, data cleaning, dealt with missing records, purged and consolidated data** for analysis and selected the appropriate visualization techniques.
- Performed **econometric analysis** for ISA investment fund utilizing python pandas, numpy and matplotlib.
- **Engineered features** and implemented hybrid **forecasting simulation models** to predict customer lifetime value (CLV), individualized risk analysis and personalized pricing.
- Created and maintained **SQL Queries & Scripts** for data integrity and database maintenance.
- Developed reports/dashboards using dual axes charts, histograms, filled map, bubble chart, bar chart, line chart, tree map, Box and whisker plot, stacked bar using **Tableau**.
- Developed an **analytical dashboard tool** using **V/H LOOKUP, VBA Macros, and custom functions** in Excel to calculate financial metrics and benchmarks to help analyze candidate risk profile and measure **KPI's** which helped to reduce entire process time by 30%.

Data Analyst, TATA Consultancy Services (Pune, MH) January 2017-July 2017
Tools: SQL, C#, HTML/CSS, Microsoft Visual studio

- Developed database schema for capturing the status of each inter-bank transfer transaction.
- Prepared SQL queries (aggregation, nested joins) for analytical reporting, provided **SQL** scripts to rectify records.
- Automated some process using **Excel Macros** to reduce time by 70%.

DATA SCIENCE PROJECTS

Data Analytics for comparing 2 investment opportunities. (Link)

- Gathered and process cashflow data for 2 investments and a benchmark index (interest free rate).
- Performed statistical analysis of both investments using Sharpe Ratio in python and interpreted results of analysis.

Python google maps API to extract location details.

- Built a python module to gather location data related to any entity listed on google maps utilizing web scraping.
- This module currently generates csv file with attributes like address, latitude & longitude values, and type of property. Very useful to perform real estate **valuation analysis**.

Future Transaction Classification Prediction Kaggle Challenge

- Develop a **LGBM ML model** for predicting customer transaction probability using 200 anonymous features provided by Santander bank.
- I built this model with different types of **regularizations** (L1 and L2), utilizing Python libraries such as NumPy, LGBM and matplotlib. I ranked in top **10%** teams in this Kaggle competition.

Toxic Comment Classification (Natural Language Processing)

- Built a multi class classification logistic and NN model that can classify comments based on different types of toxicity
- (like threats, obscenity, insult and identity-based hate) in a comment with an accuracy of 97.90%.
- Performed text analytics (Word cloud, TFIDF, feature engineering) to understand the nature of dataset.