Maheen Naqvi

Plainview, NY | maheenabbas99a@gmail.com
LinkedIN: https://www.linkedin.com/in/maheen-naqvi | **Github**: github.com/maheennaqvi
Site: maheennaqvi.github.io

EDUCATION

Queens College, City University of New York Overall GPA: 3.42
B.A. Computer Science Expected Graduation: Spring 2023

EXPERIENCE

McGill Centre for the Convergence of Health and Economics – *Data* Science *Research Intern* May 2021 – Present

- Part of the data science team conducting research on convergence-based solutions to analyze the disruptions caused by the COVID-19 pandemic.
- Independently led research into indexing and clustering a 2.3B row table for faster execution and querying.
- Devised plan and iterations for indexing. Sent final iteration for review to Database Administrator / SWE.
- Indexed table led to a 3X speed up in planning time and 2.4X speed up in execution time.
- Wrote pgSQL and R code to analyze tables with 2B+ rows and piped them into Tableau Prep to enable researchers.

Code ReSolve – *Junior Vice President* August 2021 – Present

- Develop numerous workshops on machine learning and data science.
- Hold presentations for club members and contribute to the Code ReSolve code base.

RippleMatch – *Leadership Development Intern* August 2021 – December 2021

- Selected from a pool of thousands of candidates to work closely with RippleMatch's Leadership Team.
- Leveraged various growth strategies and tools including social media, email marketing, presentations, peer and faculty member networking, to grow userbase and awareness on campus.
- Strategically assessed growth and performance metrics to improve, change and/or help design new growth strategies.

Westminister International School – *Teaching Assistant* March 2018 – June 2018

- Assisted creating syllabus and plans for Grade 1 & 2 students.

SELECTED PROJECTS

Customer Segmentation & Trend Analysis with K-Means Clustering & UMAP

- Analyzed customer trends and created a user-item matrix by aggregating quantity purchased by cluster and product attributes. Normalized data and used a K-Means Scree Plot and UMAP (high performing dimensional reduction) to visualize customer segments. Implemented interactivity to enable exploration of the data.

Time Series Modeling & Analysis

- Utilized a non-parametric smoother function (LOESS) at scale and mapped it back to all categories using tidy, purrr, and broom frameworks. Performed time series aggregations, measured seasonal changes, lags, percent differences, and cumulative calculations. Analyzed trends in the time series data by calculating a quarterly rolling mean using the zoo package.

Feature Engineering & Text Analysis

- Created numeric binary flags off various columns of data using vectorized string detection. Used R Factors to manipulate and organize categorical data into separate plots for visualization and investigation.

Sales Analysis Case Study

- Imported excel files into Rstudio, developed data model and wrangled data using dplyr and tidyr. Cleaned the raw data and detected missing values. Analyzed sales over time, created summary statistics and developed three visualizations: aggregated sales by year, product categories using ggplot2, and box plot outlier detection labeling using ggrepel.

Data Visualizations

- Developed effective visualizations using ggplot2 to determine top customers using aggregate percentages and heatmaps for purchasing habits and other multidimensional data.

TECHNICAL SKILLS

Programming Languages: T-SQL, R, Java, C++, pgSQL, Python, HTML, CSS, PHP

Packages: tidyverse, tidyr, stats, lubridate, stringr, forcats, purrr, UMAP, dplyr, tidyquant, ggplot, ggplot2, broom, ggrepel, zoo, readr, readxl, writexl, Pandas, Matplotlib

Tools: Docker, Azure Data Studio, Microsoft SQL Server, Erwin DM, VS Code, Rstudio, Tableau Prep, pgAdmin, CLion, Eclipse, Microsoft Excel, DBeaver, MongoDB, Redgate

HONORS AND AWARDS

Dean's List Fall 2021, Fall 2020
High Achievers Metric Coalition – Gold Medal Recipient 2015