

KAVIN SHAH

8579996623

shah.kav@northeastern.edu|<https://kavshah.github.io/> |<https://www.linkedin.com/in/Kavin-shah>

EDUCATION

Northeastern University, Master of Science in Information Systems, Boston

Sep 2018 - Dec 2020

Courses: Application in Engineering Development, Data Management and Database Design, Data Science Eng. Methods, Big Data Arch. and Governance, Data Warehousing and Business Intelligence, Digital Marketing Analytics

Mumbai University, Bachelor of Engineering in Information Technology, Mumbai, India

Jul 2013 - Aug 2017

Courses: OOPM, DBMS, ADBMS, Big Data Analytics

PROFESSIONAL EXPERIENCE

Data Reporting Analyst (Arizona Department of Education)

Apr 2021 –Present

- Currently working to create complex queries that are efficient and importing the data in **Power BI** to create dashboards.
- Queried and extracted data for the past 10 years and analyzed the data to find underlying reasons for having ethnic disparity in the teacher student ratio and answered the stakeholder questions by creating visualization story that **effectively communicated** the forecast and trends that would help to achieve a better teacher-student ethnic compatibility.
- Analyzing the impact of ethnicity in teacher evaluation score by tying it to the ethnicity of student and their performance, so as to understand ways to efficiently improve the student performance.

Data Assistant (Northeastern University)

Aug 2020 – Dec 2020

- Led designing a data warehouse** for new Career portal for Northeastern University COE. The data warehouse was integrated with Power BI to give weekly reports and decrease processing time by 20%.
- Handled the **ambiguous data issue** of having conflicting Visa Status and Country of permanent residence, the issue here was that the students marked as US resident were paired with different advisors and had a separate set of rules to get Co-op and so the students with wrong status had to be shifted mid semester which resulted in rearranging of pairings with advisors, extra documentations and confusion. Initially, I decided to create separate reports using SQL of students having conflicting visa status and country of residence. After presenting and discussing it with my coop director, I created certain rules that allowed us to verify students before pairing and hence reduce the issue rate by more than 90%.
- Created dashboard that highlighted the companies that reduced or stopped job postings in the previous 5 years which helped the coordinators directly reach out to those companies, eventually resulted in an increase of 13% in job postings.

Graduate Teaching Assistant – Big Data Architecture and Governance (Northeastern University)

Aug 2019 – Dec 2019

- Mentored students** to take a purposeful approach on how to design, supervise and execute data directed projects and comprehend the Big Data technology and architecture.

Database Developer and Analyst Co-Op (Granite Telecommunications)

Aug 2019 – Dec 2019

- Built the Data pipeline** for classification model utilizing pyspark library saving **25% of processing time** and team was awarded **star team of the month**.
- Constructed Machine Learning** models to identify customers with different spending patterns and trends and leveraged **Google Analytics** to market product to potential customers with different **A/B testing strategies** increasing customer interaction rate by 27%.

Data Analyst (TeknoPoint)

May 2017 – Jun 2018

- Maintained a MySQL database for storing data and created efficient queries making report generation efficient by 15%.
- Automated a report for the sales team using filters, parameters and calculated sets that automatically qualified sales lead resulting in a lift in **conversion rate of 21%**.

TECHNICAL SKILLS

BI/Reporting Tools:

Tableau, Advanced MS Excel (arrays, pivots, lookups), Power BI, Looker

Database:

MySQL, Oracle SQL Developer, SQL Server, PostgreSQL, Hive, Google BigQuery, IBM DB2

Programming languages:

Python (Pandas, Numpy, Matplotlib, Seaborn, Scipy, Scikit-learn), SQL

Statistical Techniques:

Regression Analysis, Bayesian Modeling, Time Series Forecasting, Monte Carlo Simulation

Certification:

Tableau Desktop Specialist 2020, IBM Data Science 2020, Oracle 11g SQL Certificate 2016.

ACADEMIC PROJECTS

IMDB Data and Analysis

Mar 2020 – Apr 2020

- Supervised Data Profiling using **Alteryx** to provide specific statistics and synthesized a data warehouse from multiple data sources by utilizing **Talend** and learnt how to create a one-click execution process that ingested the Database with more than 250 Million rows in 60 mins.
- Learning outcome** – Learnt the process of ingesting data using data integration tool, ETL process before loading the data in BI tool and visualizing, Connection process from Database to Talend to BI tool and the issues that can surface during the connection.

IBM Data Science Capstone

May 2020 – Aug 2020

- Applied k-means clustering along with **Foursquare** API to segment and cluster neighborhoods in city of LA.
- Ranked Neighborhoods based on income per capita, population density and found top 3 most visited places in neighborhood, aiding in recognizing where a Fast-food restaurant has potential of being more than 90% (CI) successful.
- Learning outcome** – This project was a culmination of a 9-course part wherein I learnt working in IBM Watson environment, using SQL in Python, how to use API call in Python, process of identifying data points, Clustering techniques, Visualizing clusters using Folium and working with real life data and issues.