KAVIN SHAH

Boston, MA, 8579996623

shah.kav@northeastern.edu|https://github.com/KavShah |https://www.linkedin.com/in/Kavin-shah

EDUCATION

Northeastern University, Master of Science in Information Systems, Boston, (3.45 GPA)   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 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%.
- Worked to automate and streamline existing manual reports supplementing **Power BI**, MYSQL and Python to reduce overall **processing time by more than 50%.**
- Collected Data form programming survey Qualtrics and maintained the quality of data and refreshed the dashboard weekly with the updated data.

Graduate Teaching Assistant – Big Data Architecture and Governance (Northeastern University)   Jan 2020 – Apr 2020

- **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.
- Conducted sessions to help students develop and organize a data driven enterprise to comprehend complex nature of devising and **managing data-controlled projects**.

Data Analyst Intern (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 s**tar team of the month.**
- Consolidated SQL and Tableau to develop interactive dashboards using an ODBC connection to get insights into existing and prospective customers, helping finance team to **increase sales by 10% in OHIO.**
- Migrated data from SQL Server & Oracle to **Hadoop cluster Impala** using **Sqoop** increasing execution speed **by 35%.**
- **Constructed Machine Learning** models to identify customers with different spending patterns and trends and leveraged Power BI and **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 built a view making query and report generation efficient by 15%.
- Designed reports in **Tableau** and Excel (using **VLOOKUP** and Pivot tables) for reporting and **Python** for data analysis and cleaning, decreasing decision-making process by 33%.

TECHNICAL SKILLS

BI/Reporting Tools:	Tableau, Advanced MS Excel (macros, index, arrays, pivots, lookups), Alteryx, Power BI, Looker
Database:	MySQL, Oracle SQL Developer, SQL Server, PostgreSQL, Hive, Sqoop, AWS Redshift, IBM DB2
Programming languages:	Python (Pandas, Numpy, Matplotlib, Seaborn, Scipy, Scikit-learn), SQL, R
Tools:	Talend, SSIS, Google Analytics, MS Office Suite, Google Cloud Platform, AzureML studio, Snowflake
Statistical Techniques:	Regression Analysis, Time Series Forecasting, Monte Carlo Simulation, A/B Testing
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 created a one-click execution ingesting the Database with more than 250 Million rows in 60 mins.
- Ingested Staging tables and warehouse data in MS Sql Server and established connection with BI tools such as Looker and Tableau.

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.

Subway Marketing Analytics   Jun 2019 – Aug 2019

- Deployed flask on GCP leveraging cloud shell and updating data in real time in excel as well as in Tableau using GCP.
- Constructed Main and Special cluster profiles with TSNE and on basis of customer's input and using Logistic regression to surmise acknowledged offers by achieving an accuracy of 87%.

Boston Airbnb Prediction   Mar 2019- Apr 2019

- Predicted price of the AirBnB listings with 82% accuracy, based on amenities provided and its distance from airport Boston.
- Presented top 5 listings around selected destination in Boston by creating a metrics based on reviews and price.