Karan Gupta

Austin, TX

+1 (979)985-7541 karan_mays@tamu.edu linkedin.com/in/karan-gupta- https://karan7798z.github.io/

EDUCATION

Texas A&M University, Mays Business School — **College Station, Texas**
Master of Science in Management Information Systems | GPA: 4.0/4.0 — **May 2021**

College of Engineering Pune — **Pune, India**
Bachelor of Technology in Electronics and Telecommunication| GPA: 3.05/4.0 — **May 2016**

EXPERIENCE

HDFC Bank Ltd., **Mumbai, Maharashtra | *Data Analyst*** — **July 2016 – June 2019**

- Designed **optimized queries and procedures in MSSQL** to improve the response time of digital loan webforms by 30%; elevated customer satisfaction index of loans to 97%
- Successfully interpreted key metrics for deriving transactional summary from digital loan data that contributed to raising the digital loan business of the bank by 90%; collaborated with business and marketing teams to formulate the metrics
- Architected **ETL pipelines** to integrate transactional data coming from varied data sources using Python scripts
- Developed **real-time Tableau dashboards** for stakeholders to improve visibility and analysis of daily support ticket data
- Performed **EDA** on IIS web traffic log of the primary website of the bank using pandas; provided insights which helped in reducing technical declines by 80%

Tata Consultancy Services, **Pune, Maharashtra | *Summer Analyst*** — **May 2015 – July 2015**

- Developed procedures that aggregated data received from multiple sensors for an autonomous car project
- Developed MATLAB code for fusing sensor data

PROJECTS

Dominick's Fine Foods – Data Warehousing -

- Developed a Data Warehouse for a large enterprise dataset using Kimball's design of Conformed Data Marts
- Designed a dimensional STAR model for the data and utilized SSIS, SSAS and SSRS for ETL, cube analysis and reporting

COVID-19 Real-time interactive dashboard –

- Designed a real-time dashboard in Tableau Public, utilizing the dataset provided by John Hopkins Univ.
- Developed a pipeline in Python to scrape the data from JHU's GitHub, transform the data & load the transformed data into Sheets using Google API; automated pipeline execution using Apache Airflow

Quantifying sentimentality of songs by an artist -

- Scraped data for each track by my favorite artists from Spotify and Genius using their respective API's
- Computed a metric - Gloom Index for each track by averaging the musical positivity score and the lyrical sentiment score
- Reported the results in an interactive python dashboard using bokeh

Movie Recommendation System –

- Designed a movie recommender system using Content-based filtering and Collaborative filtering approach
- Combined recommendations generated using metadata soup of movie cast, director(s), genre(s) and keywords with recommendations generated using plot description for a more holistic content-based recommender

Prediction of Speed of Spread of COVID-19 –

- Developed predictive models to estimate the rate of spread of COVID-19 using community mobility data with 66 predictors
- Reduced dimensionality using Principal Component Analysis (PCA)
- Weighed Regression based (Linear, Ridge, Lasso) and Tree-based approaches (with bagging and boosting) to choose the best model that provided lowest test MSE and interpretability of features pivotal in affecting the rate of spread of the virus

Flight Guru –

- Developed a web app for suggesting flights and destinations to users, through analysis of 8 years' worth flight records
- Created, configured and managed a cluster of 3 MariaDB databases on an AWS EC2 instance
- Migrated data from relational to a document database (MongoDB); reduced query execution time by 20%

TECHNICAL SKILLS & CERTIFICATIONS

Certifications: Machine Learning (IBM), KPMG Data Analytics Consulting Virtual Internship, Tableau Data Scientist
Languages: Python, R, Spark (w/PySpark), SQL, HQL, HTML5, CSS3, JavaScript (d3), Shell Scripting, Bash Scripting
Tools: SSIS, SSAS & SSRS, Tableau 10.5, Power BI, Advanced Excel (VBA), RStudio, Jupyter notebook
DB & Frameworks: MSSQL, Oracle 12g, MySQL, PostgreSQL, MariaDB Galera Cluster, Google BigQuery, MongoDB, Hive, Spark, Kafka, AWS Suite (EC2, S3, Amazon RDS, Amazon Kinesis Analytics)