Karan Gupta

Austin, TX

📞 +1 (979)985-7541 ✉ karan_mays@tamu.edu in 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

Amazon Business, Austin, TX | *Business Intelligence Engineer*   **June 2021 – Present**

- Improvised the Weekly Business Review (WBR) process by curating metrics and designing dashboards in QuickSight for product features that provide stakeholders with improved visibility on performance; automated the entire WBR workflow
- Developing pre-materialized SSAS cubes to meet accelerated reporting needs of stakeholders
- Kickstarted the CR process in the BI team for a more qualitative and efficient codebase
- Democratizing data across all sources by developing optimized ETLs in Redshift and AWS Glue for precise reporting
- Working closely with Data Engineers, Data Scientists and Product Owners to curate highly impactful metrics

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
- Developed real-time Tableau dashboards for stakeholders to improve visibility and analysis of daily support ticket 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 PCA and weighed multiple regression and tree-based approaches 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

Languages: Python, R, Spark (w/PySpark), SQL, HQL, HTML5, CSS3, JavaScript (d3), Shell Scripting, Bash Scripting
Tools: SSIS, SSAS & SSRS, QuickSight, Tableau, Power BI, Advanced Excel (VBA), RStudio, Jupyter notebook
DB & Frameworks: MSSQL, Oracle 12g, MySQL, PostgreSQL, Redshift, MariaDB Galera Cluster, Google BigQuery, MongoDB, Hive, Spark, Kafka, AWS Suite (S3, Amazon RDS, Amazon Kinesis Analytics, Redshift, Glue)