Karan Manoharan

http://www.github.com/karanmrn

Email : karanmanoharann@gmail.com
Mobile : +447459254319

EDUCATION

Cardiff University — Cardiff, Wales
Master of Science in Computer Science — *Sept. 2022 – Dec. 2023*

Anna University — Chennai, India
Bachelor of Engineering in Computer Science and Engineering — *July. 2014 – July. 2018*

EXPERIENCE

Tenacium DC — London, United Kingdom
Data Engineer — *July 2023 - December 2023*

- **EV Prediction**: Built an end to end data pipeline and Machine Learning model to determine the UK's electric vehicle market and the energy consumption using Time Series Forecasting
- **ETL Pipeline**: Extracted data and stored in S3 buckets and used AWS Athena for querying and analysing
- **Forecasting**: Built a model using ARIMA and SARIMA in AWS SageMaker with an accuracy rate of 85%
- **Stakeholder Management**: Managed a team of four graduates to build dashboards and coordinated with stakeholders, resulting in the completion of the project 30% ahead of schedule

Everest — Chennai, India
Data Engineer — *Jan 2021 - Aug 2022*

- **Database Migration**: Migrated data from on premise database(OLTP) to Azure SQL database for better storage
- **Data Pipeline**: Built an ETL pipeline to extract data using Azure Data Factory and scheduling using triggers for daily injection
- **Data Modeling**: Developed debug, optimized and deployed SQL queries, stored procedures and functions, resulting in a 40% decrease in data recovery times and overhead
- **Data Transformations and Warehousing**: Optimized data transformations on Databricks using PysparkSQL and python for cleaning, filtering and joining dataframes leading to reduced compute cost and storage saving 1000$
- **Reporting and Data Visualization**: Built a real-time dashboard visualisation project, using Power BI, to track key performance indicators and monitor data accuracy, resulting in a 15% increase in operational efficiency and data-driven decision making.

Orkash Services — Gurgaon, India
Data Scientist — *Sept 2018 and Dec 2020*

- **Customer Churn**: Built Decision tree-based algorithms Random Forest and Gradient Boosting Machines and collaborated with Data Scientists and engineers to improve data quality and establish new data collection protocols
- **Twitter Sentiment Analysis**: Natural Language Processing to perform sentiment analysis on posts for political parties using BERT transformer
- **Data Warehousing**: Setup, schema design and management of Amazon Redshift for daily ETL injection from various data sources
- **AB Testing**: Conducted statistical analyses and designed experiments to evaluate the effectiveness of marketing campaigns, resulting in data-driven recommendations for improving campaign performance
- **Dashboard Reporting**: Conducted exploratory data analysis and developed visualizations that identified key trends and insights in customer data, resulting in data-driven recommendations for improving customer experience

PROJECTS

- **Spotify Data Pipeline**: An ETL(Extract, Transform, Load) pipeline, to retrieve data from Spotify using Spotify API, transform it to a desired format, and perform analytics
- **Formula 1 Championship Analytics**: To design and orchestrate a Data Pipeline using Databricks and Azure to perform analysis and visualization on Formula 1 race results
- **Recommendation System**: Music and Movie recommender systems using collaborative filtering on public datasets

TECHNICAL SKILLS

- **Languages**: Scala, Python, R, C++, SQL   **Technologies**: AWS, Azure, PowerBI, Kafka, Snowflake, Dbt
  Libraries: Libraries: Pandas, NumPy, Sci-kit learn, Matplotlib, Seaborn, Pytorch, Darts, TensorFlow