ONKAR RANE

Fort Collins, CO | (970)-231-7926 | onkrane@gmail.com | https://onkarane.com |
https://www.linkedin.com/in/onkarane/ | https://github.com/onkarane | Tableau

SUMMARY

Computer Information Systems graduate with expertise in Hadoop/Spark development, serverless architecture, task automation, data/ETL pipeline design & deployment, data visualization, and database development/maintenance.

EDUCATION

Colorado State University, Fort Collins, CO	**May 2021**
Master of Computer Information Systems	**CGPA: 3.61 / 4**

University of Mumbai, Mumbai	**May 2018**
Bachelor of Engineering in Information Technology	**CGPA: 7.1 / 10**

EXPERIENCE

Graduate Research Assistant - Colorado State University, Fort Collins, CO   Jan 2021 – May 2021
- Identified and documented research requirements by conducting virtual meetings with the faculty members of College of Business every week, gathered required data and performed detailed analyses.
- Compiled comprehensive research reports of the findings that aligned with the requirements.

Business Intelligence Intern - FynTune Solution, **Mumbai, India**   Jan 2019 – June 2019
- Leveraged available business intelligence data to formulate Tableau reports using calculations and parameters.
- Collaborated with several departments to identify end users and streamline the relational database design.
- Implemented stored procedures, and views on production database server using SQL Server Management Studio.

TECHNICAL SKILLS

- **Programming Languages:** Python, C#, SQL, Java, R
- **Database:** MS SQL Server, PostgreSQL, MySQL, MongoDB
- **Database Objects:** Views, Stored Procedures, Triggers, Certificate, Index
- **Big Data/ETL Tools:** Hadoop, Spark, Hive, Airflow, Redis, SSMS, SSIS
- **Amazon Web Services:** Lambda, CloudWatch, S3, EC2, EMR, Redshift, RDS, MWAA, VPC, IAM
- **Visualization Tools**: Tableau, Looker, Plotly Dash, Seaborn, Matplotlib, Power BI

PROJECTS

Yelp ETL Pipeline (https://github.com/onkarane/etl_pipeline)   May 2021 – May 2021
Software Stack: Python, Apache Airflow, Pyspark, Pandas, S3, Redshift, Bash, Psycopg2
- Using Apache Airflow's DAG designed an ETL Pipeline to normalize unstructured JSON data with Pyspark and upload the data to AWS S3 to be copied into the AWS Redshift cloud data warehouse.
- Developed Python script to create data warehouse tables, load the data into staging tables, upsert the data into production tables and perform data integrity checks after the update is complete.

Stock Market Data Pipeline & Dashboard (https://dash-stk.herokuapp.com)   Mar 2021 – Apr 2021
Software Stack: Python, RDS, S3, CloudWatch, Lambda, Plotly Dash, Prophet
- Designed serverless data pipeline to move stock market data from the S3 data lake to PostgreSQL on AWS RDS.
- Developed Lambda Function to fetch stock market data from Yahoo Finance API, get performance predictions from the FbProphet model on Heroku cloud, store the data into the S3 data lake and load it into PostgreSQL RDS tables.
- Designed stock market dashboard using Plotly Dash to get the real-time stock prices, quarterly balance sheet overview, price forecast for next three months, historical stock performance and deployed it on Heroku.

Starbucks Customer Segmentation (https://github.com/onkarane/segmentation)   Jan 2021 – Feb 2021
Software Stack: Python, Scikit-learn, Pandas, K-Means, Principal Component Analysis, Standard Scaling
- Performed extensive data wrangling tasks on unstructured JSON data to remove duplicate customers, impute null values, eliminate misattribution, compute the response & conversion rates, and harmonize the data.
- Identified $70,000 worth of lost revenue as a result of improper targeting of customers for special offers.
- Classified customers into three distinct clusters using K-Means to be targeted for marketing based on five years of transaction and reward data. Used Principal component analysis (PCA) for dimensionality reduction.