469-380-4817 | **Samarth Khare** | Dallas, TX
khare.samarthutd@gmail.com | linkedin.com/in/samarthkhare | github.com/samarthxxkhare | samarth-khare.com

EDUCATION

The University of Texas at Dallas May 2020
M.S., Business Analytics; Data Analytics; Data Science GPA 3.2
Devi Ahilya University, India June 2018
B.B.A./ Bachelors, Business Administration; Analytics

TECHNICAL SKILLS

Analysis Tools: Python • SQL • R • NoSQL • Advanced MS Excel • VBA • Pivot Tables • SAS • Impala • Presto • Athena
Databases: MySQL • SQL Server • MongoDB • HBase • Cassandra • Hive • S3 • Redshift • RDS • DynamoDB
Big Data Processing: Hadoop • Spark • Sqoop • Flume • Storm • Flink • Tez • EMR • Kinesis • Glue
Cloud Services: AWS
CRM and ERP: Salesforce • Microsoft Dynamics • Hubspot
Data Visualization: Tableau • Power BI • Qlik • Kibana • QuickSight
Techniques: Regression • Classification • Forecasting • Machine Learning • Business Intelligence • Data Analysis • ETL • Descriptive Analytics • EDA • Clustering • Statistical Modeling • Ensemble • Feature Engineering • Big Data • Data Structures

BUSINESS EXPERIENCE

Intuilize (Dallas, TX) January 2020 – May 2020
Data Analyst | Full-time Intern **(Tools: R, Python, Tableau)**

- Performed Sales Data Analysis using Customer/Product Segmentation while working in a team of **seven** people.
- This project contained about **25** months of data on sales activity of our client's transactions and lasted **4** months.
- Performed Data Cleaning, Exploratory Data Analysis on data about **4146** customers.
- Performed Recency Frequency Monetary (RFM) Analysis on a 5-point scale, ABC Analysis to classify customers into different clusters.
- Overlaid information regarding segmentation, presented using **Tableau** dashboards.

Alliant Infotech Limited (Indore, MP, India) May 2015 – April 2018
Data Analyst | Full-time Intern **(Tools: SQL, MS Excel, Tableau, SAS, R)**

- Extracted and refined the unstructured data and loaded it to **MySQL** to study the data easily using SQL.
- Optimized data acquisition by **20%** by collecting data from heterogeneous sources and constructed single relational database.
- Increased customer retention rate by **4%** by developing Loyalty Programs for each customer segment.
- Implemented an inventory forecasting model using moving averages analysis on **Excel** and summarized historical purchase data to help the executive team better negotiate with vendors.
- Helped regulate the demand of the client's products. Developed and presented reports to stakeholders that resulted in **14%** higher revenues and margins.

PROJECTS

Snack Foods Market Analysis using Scanner Data in SAS

- This project contains scanner data on the sales of snack foods across United States.
- Performed Customer Segmentation using RFM Analysis to classify customers.
- Tested different hypotheses like effect of advertising on switching, effect of product characteristics on loyalty, etc.

Credit Card Fraud Detection using Machine Learning in Python

- This dataset contained transactions made by credit cards in September 2013 by European cardholders.
- Most of the features contained are principal components generated by PCA to protect confidentiality of the customers.
- Performed data cleaning and built classification models to predict fraud (Class = 0/1).

Rainfall in Australia Analysis using Machine Learning in Python

- This project contained about 10 years of daily weather observations from numerous Australian weather stations.
- Trained several machine learning algorithms for Classification to predict next day's rainfall based on 23 features.
- Implemented ensemble methods and dimension reduction techniques like PCA to improve scores.

Census-Income Database Analysis using Python

- This project contained data on weighted census data extracted from the 1994 and 1995 population surveys by US Census Bureau.
- **Visualized how income is controlled by race, age and gender. Explored how these factors discriminate female role in employment.**
- Compared models on basis of accuracy scores, Sensitivity and Specificity, AIC and BIC values.
- **Recommended targeting the 6% population with income >50k for fundraising and starting campaigns for women empowerment.**

Telco Customer Churn Analysis using R

- This project centered on the behavior of telecom customers who are plausibly leaving the current subscription platform.
- Predicted most striking behavior of customers, in order to increase retention.
- Recommended cost savings of 87% by implementing predictive analytics on user churn in the telecom industry.