THARANI KUMARESAN

❑ Fort Mill, SC, USA ❑ (803) 370 5633 ❑ tharanimscs@gmail.com ❑ LinkedIn ❑ GitHub

EDUCATION

Master of Science in Computer Science (GPA: 4/4), University of North Carolina, Charlotte, NC, USA 2021-2022

Bachelor of Engineering in Electronics and Communication Eng. (GPA: 8.6/10), Anna University, Chennai, India 2007-2011

CORE COMPETENCIES

Python | PyCharm | Spyder | Jupyter Notebook| Pandas | NumPy |SK-learn | Matplotlib | TensorFlow | Keras | Spark | Oracle | SQL | MySQL | Informatica | Autosys | Tableau | JIRA | BitBucket | GIT Hub| WinCVS | GCP-Cloud Storage, SQL, BigQuery, Vertex AI, DataPrep, DataStudio| Amazon AWS-S3, EC2, Athena, Glue, Redshift, Kinesis, Quicksight

RELEVANT COURSE WORK

- Deep learning on Big Data Analytics, Knowledge Based Systems, Computer Networks, Data Structure and Algorithm, Database Systems, Knowledge Discovery in Database.
- Learned GCP Essentials, Big Query Basics, Data Science on Google Cloud in Qwiklabs.
- Completed DataCamp tutorial in SQL, Introduction to Python, PySpark, ML with PySpark, Cleaning data in python.
- Completed certification in AWS Academy Data Analytics and Python

GRADUATE PROJECTS AND WORK EXPERIENCE

NBA Player Stats — 03/2022 – 05/2022

- Analyzed the performance of certain teams and players in the season 2021/2022.
- Utilized GCP for the entire ETL process.
- Extraction - Retrieved NBA player stats from Kaggle.com and loaded into the storage bucket.
- Transformation - Cleaned and merged the dataset using DataPrep and Jupyter notebook.
- Loading - Loaded the cleaned data into storage bucket to ran queries in BigQuery and for building dashboards using Data Studio.

COVID-19 Data Lake Analysis — 04/2022 – 05/2022

- Used AWS to analyze COVID-19 World cases death dataset from an open data registry for AWS, forecasting the number of confirmed COVID cases, tests performed, and total number of deaths to take appropriate actions.
- Performed Exploratory Data Analysis (EDA) using Jupyter notebook to detect outliers and to understand different patterns in the data through visualizations.
- Loaded the dataset into Amazon S3, created an amazon SageMaker instance.
- Completed Data Preparation using Jupyter notebook by cleaning and transforming the raw data prior to processing and analysis; Used Linear regression model for predicting total cases.
- Used Polynomial features to optimize the model for reducing the root mean squared error and have done visualizations by plotting polynomial regression best fit and train data for total cases for various dates.

Impact of COVID-19 on Real Estate Analysis — 10/2021 – 12/2021

- Analyzed the housing market in Chicago and Boston States from 2019 onward.
- Done data preprocessing by extracting only the required columns from the data set and imputed the missing values with mean/median.
- Visualized the data set using seaborn in Google Colab; Used ANN model to predict the housing market by splitting the dataset into 80 percent for training and 20 percent for testing.
- Used a network with 4 hidden layers and used Adam Optimizer with learning rate of 0.001 and ran the training for 500 epochs and batch size of 32.
- For the test dataset our model had the loss of 0.0153 and mean squared error of 0.015
- After prediction made visualization by comparing the predicted and the actual pending sales, new listings, total homes sold, median sale price, off market, inventory, price drops.

Data Analyst: Client, HTC global services Ltd. — 01/2012 – 05/2013

- Charged with developing technical standards, procedures, and guidelines to govern data models; with documenting success criteria; and with tracking solution effectiveness including system performance, adoption, and other key metrics.
- Author technical documentation, architecture designs, and data flow diagrams.
- Employ Type I/Type II SCD mappings to update slowly-changing dimension tables; develop UNIX shell scripts and scheduled ETL load applying AutoSys (Job Scheduling Tool); design, develop, test, and maintain the project based on user requirements