Shashidhar Reddy

[www.linkedin.com/in/shashidhardaida](http://www.linkedin.com/in/shashidhardaida) | <https://github.com/shashidhardaida?tab=repositories> |[shashidharreddy.1007@gmail.com](mailto:shashidharreddy.1007@gmail.com) | 2147358060

**Education**

**The University of North Texas**, TX, USA

*Master’s in Data Science* Cumulative GPA: 4/4.00

# Work Experience

**The University of North Texas**, TX, US —*Research Assistant in Data Science* Jan 2020 – April 2021

**Project:** Characteristics of Information Spreading across nations

* Extracted, Processed and Analyzed the COVID-19 related data from Twitter to understand social view on COVID-19.
* Developed a python script to extract the data from Twitter API in Json Format and store the files in the project server.
* Wrote 50+ Python scripts and ran them in Linux server to retrieve the useful information from the tweets.
* Created 100+ visualizations in Python and Tableau. Sampled the data from different time periods for Time series analysis.
* Transformed data and implemented extensive feature engineering for developing sophisticated Machine Learning models.

**Technologies: Python, JSON, Tableau, MS Excel, Linux, Machine Learning**

**Diebold Nixdorf India Pvt Limited,** Mumbai, India —*Data Engineer* May 2018 – Oct 2019

**Clients: Sephora**

* Processed and analyzed the customer data using **Spark, Python, and Hive**. Automated **ETL** processes using shell script and created hive bucketing, partitioning tables
* Rectifying the errors occurred in Data Pipeline while pulling the data from Relational databases to Hadoop environment.
* Involved in frequent improvement of data pipeline by troubleshooting the daily ETL processes and deep dive into the user generated data from Retail stores in order to derive the sophisticated data models
* Developed extensive dashboards using Python to analyze the data which reduced the cost of licensing other analytical tools by 80%

**Technologies: Distributed Systems, Hadoop, Spark, PySpark, Hive, Sqoop, Linux, Tableau, MySQL, SQL**

**Cognizant Technology Solutions,** Mumbai, India — *SQL Developer* Dec 2016–May 2018

**Client:** Sanofi Pasteur

* Designed an **SQL** database integrated from different tables related to the client data with normalizations.
* Developed 200+ stored procedures, queries, views, triggers, and other database objects to support internally hosted applications.
* Analyzed the existing SQL queries and modified to improve the performance of client’s applications.
* Communicated regularly with application developers to develop new tools and systems to support growing business needs.

**Technologies: Microsoft SQL, Linux**

**Academic Projects**

**Local Search System for Wiki News (Python, Django, HTML, CSS, JavaScript, Elastic Search, MySQL)**

* Designed and developed a web portal which scraps/extracts data from Wiki News and store in local database.
* The portal provides an interface for user management (Sign up, account management), Item management (Making changes in the news data and update in database), Searching the news items (A search page like google search page).

**Prediction of Airbnb prices (Python, Tableau, Scikit Learn)**

* Developed a Linear Regression model which predicts the price of hosts in Airbnb Los Angeles.
* Performed extensive data cleaning and transformation to improve the model accuracy up to 81%. Created interactive visualizations in Tableau to understand the trends in various features that affects the price.

**Data Engineering with Python (Python, Jupyter notebook)**

* Wrote 20+ Python notebooks to implement data analysis and data processing tasks on various data sets.
* Implemented various data science concepts like data analysis, Data cleaning, Data Visualization, Simulation, Sampling, Web scrapping. Used various python libraries like pandas, NumPy, SciPy, Matplotlib, Seaborn etc.

**Retail Store Sales Management System (MySQL, Data Modelling)**

* Designed and developed a relational data base in MySQL to provide an efficient Retail store sales management system.
* The database contains various tables which stores and maintains information of customers, products, Transactions, Employees.

**Technical Skills:**

**Programming Languages** Python, R-programming, SQL, C#

**UI Technologies** JavaScript, HTML, CSS3, AngularJs

**Big Data** Data Lake,Hadoop, Pig, Hive, Kafka, Spark, HDFS, Scala, Distributed Systems

**Cloud Technologies** AWS (S3 storage, Ec2 service), Azure

**Analytical Tools** Tableau, Microsoft Excel, Splunk

**Databases** Microsoft SQL, MySQL, AWS Redshift, SAP HANA

**Machine Learning** Linear Regression, Logistic Regression, Decision Tree, Random Forest, SVM, Classification, Clustering,

KNN, PCA, Keras, Tensorflow