**Disha Neve**

San Diego,California|dishaneve@gmail.com **|** (813)606-2679|[LinkedIn|](https://www.linkedin.com/in/DishaNeve/)

**Professional Summary:**

# Experienced Data Science Engineer working on extracting and transforming data from heterogeneous data sources and playing a pivotal role in applying advance ML and analytics techniques for predicting and determining product performance.

# Experience in building efficient analytical pipelines and portraying the results using Visualization framework Plotly and Dash and tools

# like Tableau.

**Education:**

**Master of Science in Business Analytics and Information Systems** (*GPA: 3.54/4.0)*  *August 2017-Dec 2018*

University of South Florida, Tampa, FL

**Bachelor of Technology**, **Electronics and Telecommunication** * August 2011-August 2015*

Vishwakarma Institute of Technology, Pune, India

 **Skills**:

**Programming Languages**: R, Python, SAS **Databases:** SQL Server,Teradata, Oracle Server, Presto

**Data Mining Tools**: SAS Enterprise Miner, Rapid Miner, Weka**. Reporting/Visualization:** Tableau, Looker, Power BI, DASH

**Big Data Technologies**: Hive, PySpark **Cloud Technologies**: Amazon Web Services, Google Cloud Platform

# Work Experience:

# Teradata Corporation, San Diego, USA *Feb 2019-Present*

# R&D Team, Data Science Engineer

# Built advanced analytical pipelines in python for extracting and processing data from various data sources like Data Warehouse, Logstash (ELTK) for predicting and visualizing product performance metrices.

# Working as a part of Teradata Test Automation team to create analytical reports which draw deep statistical insights and displaying them in

# an interactive format using custom dashboard building tool Dash using Plotly.

# Defining new metrics for product performance evaluation, designing custom made reports and dashboards for various KPI’s based on

# business user requirements for A/B Testing.

# Worked on test metadata automated collection process using kubeless functions in python script and optimized the performance and reduced computing time by using parallel function execution.

# Working on designing a simplified, automated and robust methodology PEM (Performance Evaluation Machine) which consolidates results

# from various reports and highlights the most critical issues to the users. Second phase involves predicting the issues based on historical

# data and providing recommendations to end users.

# Providing training sessions and presentations to customers on how to use improve product decision making using advance analytical

# reports, dashboards and data inferences.

# Teradata Corporation, San Diego, USA *May 2018-August 2018*

# R&D Team, Data Engineering Intern

# Worked as a part of Teradata Test Automation team to implement new test reporting system and work to build data pipelines in Python.

* Worked on Query Level Analysis Project which helps to analyze the difference in the query performance for different version of databases launched in such a way that same query is compared across various versions of databases by an automated python script in kubeless environment.
* Anomaly detection by analyzing outliers in various parameters which helps to understand key factors affecting query behavior using **SQL** and **Python** and working on trendline query which can forecast the time required and other parameters for a particular query.
* Generated reports using whichgives the performance difference of various metrics between them using graphs. It helps to find key parameters affecting query which helps in the next version build. Used JIRA for tickets and GitHub for version control of codes.

# Cybage Software Pvt. Ltd, Pune, India  *Aug 2015-June 2017*

**Media and Entertainment, Business Intelligence Developer**

* Performed SQL query optimization which improved results by 35%, designed SSIS packages for automating ETL processes.
* Worked on Reporting tools like **Tableau, QlikView** and Power BI for data Analysis and developed user i**nteractive dashboards,** worked closely for delivering Business Intelligence solutions in a data warehouse environment. and led a team of five and helped with SQL trouble shooting.

#  Academic Projects:

# Statistical Data Mining: Rossman Store Promotion Analysis Jan 2018

* Designed panel data model for retail sales prediction and forecasting of future sales for 6 weeks for 1115 stores.
* Used **R** and checked OLS assumptions like Multicollinearity, Autocorrelation, Homoscedasticity, Normality.
* Implemented Panel data models like Pool, Random and Fixed models and compared all the models.
* Performed 4 hypothesis testing and concluded results based on R square, adjusted R square, p value and other factors. using **Oracle**.

 **Statistical Data Mining: Census Data Analysis and Employment Trends in United States:**  February 2018

* Analyzed census data and predicted whether income exceeds $50K per year by following end to end modelling process involving Exploratory Data Analysis and established hypothesis of the data and used Tableau to show Most paid job, Full -Time vs Part-Time.
* Created training and validation datasets using Stratified Random Sampling(SRS) of data. Applied classification model on training set using Logistic Regression, SVM and performed validation of the models using ROC curve, Confusion Matrix and selected the final model.

 **Data Science Programming: Self-Learning and Playing Blackjack Using Python**: March 2018

* Developed a black Jack playing game which learns by itself and decides to take a hit or stand depending upon player and dealer total.
* Trained a classification model using Neural Network (Keras) and XG Boost for accurate decision making.