

# Suraj Malpani

[surajmalpani23@gmail.com](mailto:surajmalpani23@gmail.com) | 📞: (682) 331-2509 | <https://www.linkedin.com/in/suraj-malpani> | <https://surajmalpani.github.io/>

**SUMMARY:** I am a diligent individual with professional experience of 3+ years and a master's degree in the Data Science and Analytics domain. I have experience of data related consulting in diverse teams including operations, IT, engineering, etc. My motive is to derive actionable insights from the data; insights, which will be impactful for the betterment of the businesses.

## EDUCATION

**M.S., Business Analytics (Data Science track), The University of Texas at Dallas**

Dec 2019

**B.E., Electronics and Telecom Engineering, University of Mumbai**

June 2015

## BUSINESS EXPERIENCE

**Data Analysis Prof Lead, Nike (Contract), Beaverton, Oregon**

Feb 2020 – Present

- Developed a centralized data reporting system to gather 100+ factory excess materials inventory data, while working with internal product creation teams and business stakeholders in effort to save these materials; promisingly saving ~\$40M/year.
- Technologies: Alteryx, Tableau, SQL, Python, Excel, AWS, Snowflake.

**Data Science Intern, Daimler, Portland, Oregon**

Jan 2019 – Jul 2019

- Saved up to 80% of the time/week for the compliance team for analyzing daily updated vehicle complaints data on a govt website, by developing an R Shiny app that gathers, transforms, and visualizes this data in intuitive & user-friendly manner.
- Automated the data extraction and statistical algorithmic solution using SQL, Python and Alteryx, saving up to 90% time.
- Utilized machine learning (regression, clustering) and visualizations to find patterns and reduce the manual work in assembling the modules on the truck on accurate locations for the custom orders; saved 40+ hours/week.
- Predicted Warranty claims with ~78% accuracy by developing machine learning models using vehicle's telematics data.
- Technologies: Python (numpy, pandas, scikit-learn), R, Microsoft Azure Cloud, Azure ML Studio, SQL, Power BI, Tableau.

**Teaching Assistant, The University of Texas at Dallas, Dallas, Texas**

Aug 2018 – Dec 2018

- Assisted in designing the coursework, evaluated the assignments and quizzes for courses 'Business Analytics with R', 'Database Foundations' and 'Object-oriented Programming with Python', each class consisting of approx. 60 students.

**Data Science Intern, Analytics School Online, Remote**

May 2018 – July 2018

- Introduced 2 Case-study based courses on Machine learning topics Logistic Regression & Decision trees using R, Python.

**Application Development Analyst, Accenture, Mumbai, India**

Oct 2015 – May 2017

- Efficiently contributed in all phases of the SDLC from planning, designing, developing, testing & deploying of the product which focused on mail redirection, fraud monitoring services to consumers which was forecasted to generate revenue of £10M+ per quarter, all while liaising with the client, business analysts, developers, quality assurance and rest of the team.
- Developed 100+ SQL Queries (Teradata) and 50+ ETL pipelines for acquiring, modeling & transforming the data to create 17 Products report files; Scheduled cron jobs using bash shell scripts (CLI) for generating automated data reports.
- Automated the process of manual assurance of naming conventions in DataStage ETL data pipelines by developing a 'Code Review' tool in Microsoft Excel using VBA and a UNIX script, effectively saving approx. 40 hours/month.
- Designed dashboards in QlikView for purposes ranging from ERP to presenting KPIs, insights. Debugged ETL jobs, SQL queries and bash/shell scripts among ad-hoc SDLC tasks in a Project where client has database of 90% of UK's population.
- Accomplished ad-hoc tasks utilizing Big data and BI tools, Hadoop, Apache Spark, R, Informatica, SQL and Tableau.

## ACHIEVEMENTS

**Copart Data Challenge (Machine Learning, R, Predictive Analytics)**

Nov 2018 – Dec 2018

Selected to present ML model (xgboost) which predicts prices of auctioned cars (with 90%+ accuracy) to Copart's executives.

**Texas Level R Competition, Winners 2<sup>nd</sup> Place**

Apr 2018 – Apr 2018

Analyzed New York City's crime dataset containing 5.5 million rows, segregated the data based on boroughs and analyzed crime types based on location, time of the day, employment. Visualized this geospatial data using 'ggmap', 'ggplot2' in R.

## ACADEMIC PROJECTS

**H-1B Visas Analysis (Interactive R Shiny Web Application)**

Oct 2019 – Jan 2020

**Retail Data Analysis (Marketing, Predictive, Prescriptive, Analytics, SAS)**

Sep 2019 – Dec 2019

**Tweets Classification for Disaster Identification using Spark (NLP, PySpark, MLlib)**

Dec 2019 – Dec 2019

**Marketing Solutions (IBM Cognos, Cloud, Watson, Python, Customer Segmentation)**

Oct 2019 – Dec 2019

**NFL Games Scheduling (Python, pandas, git, cplex, gurobi, docplex)**

Oct 2019 – Dec 2019

Optimized scheduling of NFL Games using docplex package (IBM Cplex) in python while fulfilling all the constraints.

**Twitter Buzz Prediction (Deep Learning, Regression, Tensorflow, Keras, Scikit-learn)**

Nov 2018 – Dec 2018

Developed GLM, Ridge, Lasso, SGD, SVM, KNN Regression models to find the best model to predict Twitter Buzz.

Applied Bagging, AdaBoost and Gradient boosting models. Used PCA to reduce features on this data which had 75+ features.

**Forecasting Car Sales (Time-series forecasting, ARIMA, R, Tableau)**

Jan 2018 – May 2018