

# 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 the diligent individual who have been working in the Data Science and Analytics domain for 5+ years. I have mastered the analytics skills by working in various groups viz., Operations, IT, Engineering, in different organizations. What drives me is to design impactful solutions and extract actionable insights for advancing business using the power of data.

## BUSINESS EXPERIENCE

### Lead Data Analyst, Nike, Beaverton, Oregon

Feb 2020 – Present

- Engineered and implemented a centralized data reporting system to collect excess inventory data from 80+ factories.
- Established the solution consisting of data pipelines to collate the monthly reported data, join it with internal cloud databases, and intuitively present it to the stakeholders in Tableau dashboards, analyzing trends and report to directors.
- Liaising with Product Creation, Material Planning, Manufacturing, Sustainability, and stakeholders to salvage the excess materials stranded in the supply chain, promisingly saving ~\$40M/year, and continuously improve the processes.
- Delivered the initial scalable solution in less than 4 months and persistently improving using clustering of materials, A/B testing, modern tech stack, and gathering feedback to improve the user experience and utilization of available materials.
- Technologies: SQL, Tableau, Python, Alteryx, Excel, AWS, Snowflake, Teradata, React.js, R Shiny.

### Data Scientist Intern, Daimler, Portland, Oregon

Jan 2019 – Jul 2019

- Pioneered a machine learning model using vehicle's telematics data which predicted warranty claims with ~78% accuracy.
- Remodeled a manual data extraction, statistical algorithmic process using SQL, Python and Alteryx, saving up to 90% time.
- Conserved 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.
- 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.
- 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

- Mentored students, Evaluated the assignments & quizzes, Assisted in curating coursework for 'Business Analytics with R', 'Database Foundations' & 'Object-oriented Programming with Python', each course enrolled by ~60 students.

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

May 2018 – July 2018

- Launched two project-based courses, 'Logistic Regression & Decision Trees using R and Python' and 'A/B Testing'.

### Application Development Analyst, Accenture, Mumbai, India

Oct 2015 – May 2017

- Engineered 100+ SQL Queries (Teradata), 50+ ETL data pipelines for acquiring, modeling & transforming the data to create 17 product reports. Scheduled cron jobs using bash shell scripts (CLI) for generating automated data reports.
- Devised and implemented a 'Code Review tool' in Microsoft Excel using VBA and a UNIX script to automate the process of manual assurance of naming conventions in DataStage ETL data pipelines, effectively saving approx. 40 hours/month.
- Contributed to all phases of the SDLC from planning, implementing, testing & deploying while leveraging agile framework (scrum) of the product which focused on mail redirection, fraud monitoring among other services to customers.
- Liaised with the client, business analysts, developers, quality assurance and rest of the team during agile sprints to efficiently releasing the product which was forecasted to generate revenue of £10M+ per quarter.
- Designed dashboards in Tableau for purposes ranging from ERP to presenting KPIs, insights. Debugged ETL jobs, SQL queries, Spark, R, bash/shell scripts among ad-hoc tasks in a project where client has database of 90% of UK's population.

## 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

## 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 having 5.5 million rows, reported the geospatial data using 'ggmap', 'ggplot' in R.

## ACADEMIC PROJECTS

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

Oct 2019 – Jan 2020

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

Dec 2019 – Jan 2020

### 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 satisfying all 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.