

# SHWETA ANIL ABHYANKAR

## Data Analyst | Data Scientist

Santa Clara, CA | [shweta.abhyankar0303@gmail.com](mailto:shweta.abhyankar0303@gmail.com) | (201) 774-8040  
<https://www.linkedin.com/in/shweta03/> | <https://github.com/shwetaabhyankar>

### PROFILE

Thorough and meticulous data scientist passionate about helping businesses succeed. Proficiency in predictive modeling using machine learning and data visualization by telling persuasive data stories to deliver key insights.

### EDUCATION

**Master of Science, Analytics, Harrisburg University, Pennsylvania**

**July 2021 - December 2023**

**Master of Science, Industrial & Systems Engineering, State University of NY, Binghamton**

**January 2017 - May 2018**

**Bachelor of Science, Instrumentation Engineering, University of Mumbai, India**

**August 2008 - May 2012**

**Data Science Fellow - Women in Data (Project: EnerDapt), New York Cohort, USA**

**April 2019 - May 2019**

- Engineered a web scraper to extract property listing data from 5 different sources.
- Collaborated with a team of 7 to create a Logistic Regression model to classify and rank real-estate properties.

### TECHNICAL SKILLS

- Programming Languages - Python, SQL, R (basic).
- Machine Learning and Big Data - Python (scikit, pandas, numpy, matplotlib), Spark (basic), Hadoop (basic).
- Data Science and Miscellaneous Technologies - A/B testing, ETL, Data Science Pipeline (cleaning, wrangling, visualization, modeling, interpretation), Statistics, Time Series, Experimental Design, Hypothesis Testing, OOP, API, Excel, PowerPoint.

### MANAGEMENT SKILLS

- Collaboration and Teamwork, Problem Solving, Critical Thinking, Business Process Improvement.

### PROFESSIONAL EXPERIENCE

**Data Scientist – Wave Solutions Inc. (Client: Microsoft), Bellevue, WA, USA**

**March 2020 - April 2022**

- Analyzed market campaign effectiveness using difference-in-difference, BSTS, and Propensity Score Matching to make recommendations that helped increase YoY sales growth by 15%.
- Implemented time series anomaly detection using Python and Kensho on 2TB data to reduce analyst time by 20 hours/week.
- Developed gaming sector customer segmentation using K-Means clustering to identify 5 clusters used for marketing planning.
- Created dashboards with key performance indicators (Revenue, Orders, Visits) for business in Adobe Analytics and PowerBI.

**Data Scientist – Veracity Software Inc. (Client: Avangrid), Orange, CT, USA**

**July 2019 - February 2020**

- Collaborated with data engineers to implement ETL process and extracted data using SQL for analytical requirements.
- Built a POC model to improve transformer smart meter connection using correlation analysis with an accuracy of 95%.
- Predicted time to failure of a transformer by performing EDA and Feature Engineering to build a regression model with 30% better R-square than baseline.

**Data Science Intern - Live in Bing (Fortress Management LLC), Binghamton, NY, USA**

**July 2018 - March 2019**

- Web scraped data using Python and Selenium browser reducing data collection time from 6 hours to 2 hours.
- Automated data pre-processing by creating a data pipeline saving 10 man-hours/week.
- Executed 3 regression models to predict housing prices with Random Forest model resulting in lowest MAPE of 12%.

**Operations Analyst - Digital Media Solution, Mumbai, India**

**February 2013 - December 2016**

- Conducted quantitative analysis by using surveys to recommend pricing decisions that increased revenue by 8%.
- Inspected trends in demand, supply, and inventory and created forecasts, improving inventory accuracy by 12%.
- Designed dashboards in Excel using pivot tables and Vlookups to monitor KPIs, increasing monthly profits by \$5000.

**Intern - Digital Media Solution, Mumbai, India**

**June 2012 - January 2013**

- Gathered and identified key entities and relations to develop a customer database using MySQL.
- Interfaced with customers to receive valuable product insights to initiate future product development.

### DATA SCIENCE AND ANALYTICS PROJECTS

#### Hospital Readmission Prediction

- Identified factors that led to high readmission rate of diabetic patients within 30 days post discharge and predicted the high-risk diabetic patients who are likely to be readmitted.
- Conducted Exploratory Data Analysis which involved cleaning data, exploring features and the correlation between features.
- Worked on models like Logistic Regression, Random Forest Classifier, and Light GBM Classifier to predict readmission. Light GBM Classifier resulted in Recall Rate between 60% to 65%.