

# Shubham Khode

Chicago, IL | (858) 305-7003 | [skhode2@uic.edu](mailto:skhode2@uic.edu) | [linkedin.com/in/shubhamkhode](https://www.linkedin.com/in/shubhamkhode) | [www.shubhamkhode.me](http://www.shubhamkhode.me)

## EDUCATION

### University of Illinois Chicago

Master of Science in Business Analytics (GPA: 4.0/4.0)

August 2021 – December 2022

### Veermata Jijabai Technological Institute (VJTI), Mumbai, India

Bachelor of Technology in Electronics & Telecommunications Engineering (GPA: 7.5/10)

August 2012 – April 2016

## SKILLS

- **Technology:** Python (NumPy, Pandas, Sklearn), SQL, R, C#, Hadoop, Spark, AWS (EC2, S3, Redshift), Git, Tableau, Power BI, Excel
- **Core Competencies:** Data Mining, Machine Learning, Predictive Modeling, Descriptive & Inferential Statistics, Feature Engineering, Exploratory Data Analysis, Time-Series Forecasting, A/B Testing, Natural Language Processing (NLP), ETL, Visualization, Jira, Agile

## PROFESSIONAL EXPERIENCE

### Chicago Transit Authority | Data Science Analyst Intern | Chicago, IL

September 2022 – Present

- Extracted 5+ years of asset maintenance data using Redshift and performed an in-depth analysis to spot key trends & anomalies
- Developed statistical risk models using Prophet, ARIMA, & XGBoost for predicting the cost of asset failures and its impact on ridership; applied multiprocessing and grid search techniques to observe forecasts with MAPE of 1.76
- Assigned risk scores for maintaining assets over time and created decision-support dashboards in Power BI for business heads to optimally prioritize and allocate ~\$100M of asset rehabilitation funds

### Wipro Limited | Data Analyst | Vienna, Austria

August 2019 – June 2021

- Conducted data analysis on customer profiles as well as their products, services, and upgrade cycles to discover underlying patterns for mitigating churn rate; made recommendations to bump up retention by 15%
- Applied A/B testing on different variants of landing/product pages for optimizing client's online shop; lifted click-through probability and conversion rate by 7% with a marked increase of ~\$5M in revenue flow
- Designed ETL pipelines to fetch and manipulate data from disparate transaction database systems and consolidate it into a data warehouse for downstream BI reporting & analytics
- Identified borderline credit risk customers using ensemble ML model; redesigned order workflows resulting in 2x faster lead time
- Built interactive Tableau dashboards for quick, actionable insights on core business KPIs; saved 8 hours/week in manual reporting
- Collaborated with key stakeholders and leveraged data-driven opportunities for aligning product roadmap to strategic goals

### Wipro Limited | Software Engineer | Pune, India

August 2016 – July 2019

- Analyzed consumer behavior and utilized prescriptive analysis for fine-tuning the portfolio of a leading telco client based on levers such as voice/data volume, speed, add-ons, and promotions; observed ARPU uplift of ~3%
- Implemented complex SQL scripts for applying pseudonymization as per EU GDPR compliance; transformed millions of customer records with sensitive information using data masking procedures
- Automated log management with an integrated framework developed using Python to monitor application activity and performance; improved bug resolution time on production by 35%, saving ~\$500K/year
- Gathered business requirements, documented high-level technical specifications, and designed E2E solutions while ensuring feasibility & integration with legacy systems
- Developed and enhanced 5+ business-critical web applications in .NET to provide a unified experience for users across platforms

## PROJECTS

### Fine-grained, Scaled Forecasting of Walmart's Retail Sales (Python, Spark, Forecasting)

- Analyzed influence of factors like markdowns, holidays, fuel price, temperature, etc. on weekly sales data with 450K+ observations
- Built hundreds of time-series models in parallel by leveraging distributed data processing in Spark and generated precise forecasts for each store-department combination, achieving an 87% reduction in time for training collection of models

### Loan Default Prediction and Investment Decisions for Lending Club (R, EDA, Classification)

- Performed univariate, multivariate, and correlation analysis to identify key characteristics and driving features behind loan defaults
- Trained models like Decision Tree, Random Forest, and GBM for predicting probability of loan default as well as estimating expected annual returns; evaluated performance using AUC, ROC, and lift curves and combined best models that maximized ROI

### Market Segmentation of Soap Company (R, Clustering)

- Executed K-means, K-medoids, and Hierarchical clustering techniques for profiling customers based on their buying behavior and basis of purchase and designed more effective campaigns with targeted promotions

### Spotify – Popularity Analysis and Recommendation (Python, Tableau, Regression)

- Investigated the effect of acoustic features such as energy, danceability, valence, etc. across millions of top-charts songs and derived insights to build song popularity prediction regressors and recommendation systems