

Soham Bhagwat

sohambhagwat2@gmail.com | [linkedin.com/in/soham-bhagwat](https://www.linkedin.com/in/soham-bhagwat) | 812-327-4070

EDUCATION

Indiana University, Bloomington May 2024
MS in Data Science

Savitribai Phule Pune University, Pune May 2022
B.S., Computer Engineering (Honors in ML)

EXPERIENCE

Data Analysis & Modelling Co-Instructor, Indiana University Bloomington August 2023 – May 2024

- Conducted bi-weekly training for 40+ graduate students in advanced **regression analysis** and **statistical concepts**.
- Delivered instruction in **predictive analytics**, **hypothesis testing**, and **data visualization** with ggplot2 in R.
- Facilitated weekly office hours and received an average rating of 4.8/5 and positive feedback from over 90% of students.

DevOps Engineer Intern, PTC Inc. - Pune, India August 2021 – June 2022

- Automated Windchill build reports using **Python** and **bash** scripts saving 3+ manual hours daily.
- Managed migration of on-prem **SQL** Server database to **Azure SQL** DB for more than 1TB of data daily achieving 40% surge in migration speeds
- Reduced test execution time by 30% for 100k+ tests by optimizing Windchill **CI/CD** pipelines using multithreading.
- Streamlined **Oracle DB** deployment using Terraform reducing manual deployment time by 2+ hours daily.
- Developed **Ansible** scripts for automatic deprovisioning of **Azure** resources saving more than \$24000 annually

PROJECTS

NYC School Bus Delays Analysis

- Engineered robust **ETL** data pipelines using **Airflow** for 600k+ NYC school bus delay records, storing raw data in **AWS S3**
- Transformed raw data using **Python** scripts within **Airflow** and stored processed data in **PostgreSQL** data warehouse, realizing 45% reduction in data processing time
- Optimized **Power BI** dashboards for visualizing trends with auto-refresh achieving 30% reduction in data retrieval time

Chicago Traffic Crashes Dashboard

- Analyzed over 800k traffic crashes from Chicago and preprocessed data using **Excel** and **Power Query**
- Transferred the data into **MySQL** database using **Python** scripts and optimized analytic queries to improve data retrieval efficiency by 30%
- Proposed the recommendations using **Power BI** dashboards, prompting actions to reduce fatalities by over 40%

New York State Artists' Income Survey Analysis:

- Collaborated with [CCRNY](https://www.ccrny.org/) to analyze survey data of more than 25000 artists in the state of New York.
- Wrote **Python** scripts to validate, clean and preprocess the survey data, reducing process time by 50%
- Designed 5+ **Tableau** dashboards to present insights to non-profit leaders, aiming to raise \$100k for underpaid artists.

Reddit Hate speech Detection Bot:

- Developed ML based Hate speech classification bot using **SVM** algorithm, trained on over 100k social media posts.
- Programmed the bot as a **Python** script and deployed it using **AWS Lambda**
- Accurately classified and flagged 90% of hateful posts, reducing manual efforts by 100%

SKILLS

Python, R, Java, MySQL, PostgreSQL, SQLite, BigQuery, Github, Github Actions, Hadoop, Tableau, Power BI, Looker Studio, Airflow, Git, Jenkins, Ansible, Excel, Power Query, AWS, GCP, Github Actions, Mage AI, Scikit-learn, Scipy, PyTorch, NLTK, dplyr, Pandas, Numpy, Streamlit