

Kanika Chopra

kanikadatt@gmail.com | github.com/kanikadchopra | kanikadchopra.com

// Skills

Languages: Python, R, SQL, Presto, MATLAB

Tools: Airflow, Tableau, SQLServer, Hive SQL, Treasure Data, Piper

Libraries: pandas, NumPy, Scikit-Learn, nltk, genism, matplotlib, Plot.ly, OpenCV, StatsModels, GGPlot, Tidyverse

// Education

University of Toronto

Master of Science, Statistics

Sept. 2022 – April 2023

University of Waterloo

Bachelor of Mathematics, Statistics [GPA: 88%]

Sept. 2017 – April 2022

Relevant Courses: Forecasting, Statistical Learning – Classification, Experimental Design, Generalized Linear Models, Data Visualization, Data Types and Structures

// Experience

Uber – Data Analyst (Safety & Insurance)

May 2021 – July 2021

Hive SQL, Piper, Tableau

- Developed data pipelines to measure safety product metrics and designed a Tableau dashboard for monitoring performance
- Collaborated with product managers, data scientists and engineers to understand the use cases and present metric read-outs
- Conducted deep dive investigations into driver vehicle data to identify potential business opportunities

Wish – Data Analyst/Scientist (Data & Relevancy)

Jan. 2021 – Apr. 2021

Python, Airflow, Hive SQL, Treasure Data, Pandas, NumPy, StatsModels

- Built four complex data-pipelines to productionize a classification model, decreasing time-to-run by 75% using *Airflow*
- Developed a deterministic shipping time model and *Holt-Winters Additive* forecasting model to predict future merchant cash flows
- Devised a deterministic model to predict empty packages shipped by merchants to decrease fraudulent activity with 95% accuracy

Intact DataLab – Data Scientist

May – Aug. 2020

Python, Pandas, NumPy, Scikit-Learn, OpenCV, Matplotlib

- Conducted deep dive investigations on model predictions for 420K images to find improvement strategies
- Automated the development of a multi-page analysis spreadsheet to display model's accuracy decreasing time-to-create by 95%
- Researched and implemented image augmentation techniques using *OpenCV* to improve classification by 3%

RBC – Strategic Initiatives Analyst

Jan. – Apr. 2019

Tableau, SQL, Selenium, BeautifulSoup

- Automated reporting by calculating metrics using data in SQL databases to display in Tableau dashboards for business decisions
- Visualized event registration data with a *Tableau* storyboard for marketing to display a demographic and geographic analysis

// Research

- **Research Assistant (University of Toronto):** Researching fairness in machine learning to publish a tutorial paper with simulations and healthcare data analysis for clinicians. Under the supervision of Dr. Jessica Gronsbell.
- **Research Assistant (University of Waterloo):** Developed a plotting tool, [projplot](#), to assist users with testing optimization when building optimizers by providing additional visualizations. Under the supervision of Dr. Martin Lysy.

// Projects

- **[Highlighting Ethnic Biases in COVID-19:](#)** Leveraged word embeddings to quantify the biases towards Asians from 140K global articles surrounding COVID-19
- **[Predicting Parkinson's Disease:](#)** Built a logistic regression model to predict PD using auditory speech signals with 85% accuracy
- **[Netflix Browsing Time Experiment:](#)** Conducted an experiment to determine the optimal combination of preview length, match score and tile size to minimize average browsing time spent on the homepage

// Leadership

- **Shad (Program Assistant):** Taught students how to leverage Python for Data Science as part of a month-long enrichment camp.
- **Tech+ (Co-Director):** Led a team of 40 students to discuss and advocate for diversity, equity, and inclusion in Tech.
- **UW Data Science (President):** Co-hosted a *Clustering for Image Analysis* workshop for high school students interested in STEM. Organized a guided project to help 100+ students build a data science side project and network with data scientists