Kevin Dang

(647) 909-9810 | dang.kevin@outlook.com | dang-kevin | dang-kevin

SKILLS

Languages Python, R, SQL, Stata, HTML, CSS, MATLAB
Software/Tools Access, Excel, Tableau, Github, ArcGIS, LaTeX

EXPERIENCE

Data Analyst

University of Toronto - Centre for Industrial Relations and Human Resources *Sep. 2019 - Present*

- Spearheaded a data analysis project about agriculture worker wages with **Dr. Greg Distelhorst** and a team of researchers
- Performed **data cleaning** and **data wrangling** on millions of rows of data using the **dplyr**, **tidyr** and **lubridate** libraries
- Visualized relationships between variables and time trends with **ggplot2** using technical details and explanations geared towards a non-technical audience (C.F.O. of a U.S. company)

Statistical Consultant

University of Toronto - Department of Statistical Sciences *Sep. 2019 - Apr. 2020*

- Wrote a report containing data analysis work for a management consulting firm to answer business problems related to improving the health of users of their platform
- Completed a data analysis project to determine **the effect of auditory distraction on cognitive flexibility** for university students and made recommendations to improve study habits

Research Assistant

Rotman School of Management *May 2018 - Aug. 2019*

- Worked with **Dr. Chris Liu** and a team of graduate students on projects about scientific publications and careers
- Queried scientific databases using **Python-based API-Wrappers**, worked with **Pandas** dataframes and exported data for a foreign collaborator
- Merged and manipulated large datasets with **Stata**, extracted desired information, cleaned data and generated new variables
- Used **BeautifulSoup** for web scraping and exported data into Excel to improve efficiency in creating new datasets

PROJECTS

The Impact of an Earnings Call on a Firm's Share Price

- Performed sentiment analysis using using **TF-IDF** to create a custom dictionary of sentiment scores based on the training data of quarterly earnings call transcripts and created data visualizations using **seaborn** and **matplotlib**
- Used sentiment scores and word count as model inputs, obtained mean squared errors of 0.000774 and 0.001384 using **k-nearest neighbours** and **XGBoost** algorithms respectively

Predicting Credit Card Approvals

- Cleaned Pandas dataframes by filling in missing values with **mean imputation** or most frequent observations, used **label encoding** to convert non-numeric data to numeric format and split data into train and test sets
- Scaled features, fit a **logistic regression classifier** using scikit-learn with **84% accuracy** and performed a **grid search** of the model parameters to improve the model's ability to predict credit card approvals

Degrees That Pay You Back

- Cleaned data and used **elbow**, **silhouette**, and **gap statistic** methods to determine the optimal number of clusters to be used in applying the **k-means algorithm** to the data
- Visualized starting and median salaries with **ggplot2**, plotted clusters to look for patterns in career growth for certain majors

EDUCATION

University of Toronto

Honours Bachelor of Science *Sep. 2016 - Apr. 2020 (expected)*

- **Applied Statistics Specialist, Mathematics Minor**
- Courses: Computer Programming, Data Analysis, Design & Analysis of Experiments, Geographic Information & Mapping, Machine Learning, Statistical Computation, Statistical Consultation & Collaboration, Time Series Analysis