Kevin Dang

Data Analyst & Statistical Consultant

📞 (647) 909-9810 | ✉ dang.kevin@outlook.com | ⌨ dang-kevin | in dang-kevin

Education

University of Toronto *Toronto, ON*
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

Skills

Programming Python, R, SQL, Stata, MATLAB, HTML, CSS
Other ArcGIS, LaTeX, MS Access, MS Excel

Experience

Data Analyst *Toronto, ON*
University of Toronto - Centre for Industrial Relations and Human Resources *Sep. 2019 - PRESENT*

- **Spearheaded a new data analysis project** about agriculture and worker rights 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** with specific details geared towards non-technical audiences

Statistical Consultant *Toronto, ON*
University of Toronto - Department of Statistical Sciences *Sep. 2019 - PRESENT*

- Currently working on a **data analysis project** for a management consulting firm to answer their business problems
- Previously completed a data analysis project to determine **the effect of auditory distraction on cognitive flexibility** for university students

Research Assistant *Toronto, ON*
Rotman School of Management *May 2018 - Aug. 2019*

- Worked under the supervision of **Dr. Chris Liu** with a team of graduate students on projects about scientific publications and careers
- **Queried scientific databases using Python-based API-Wrappers**, worked with **dataframes** using **Pandas** and exported data into csv files
- **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

Brand Ambassador *Toronto, ON*
Mosaic North America *Oct. 2015 – Sep. 2017*

- Promoted different types of brands for numerous companies and **consistently increased product sales** by more than the daily target of **25%**
- **Drafted reports** containing information regarding customer interaction, sales made, products purchased, demo issues and conflict resolution

Projects

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 the starting and median salaries with **ggplot2**, plotted each cluster individually to look for patterns in career growth for certain majors

Super Bowl, T.V. & Halftime Shows

- Investigated tables containing Super Bowl, television, and halftime show data by **writing queries** containing various **filter and join clauses**
- Explored questions involving game outcomes, T.V. viewership, ad costs and musician performances

Radius of the Earth

- Collected data on gravitational strength using a gravimeter, manipulated data with **NumPy** and fit **linear regression** models using **SciPy**
- Plotted models with **Matplotlib**, performed **chi-squared analysis** and estimated the Earth's Radius to **within 30 kilometres**