

Jack Li

Data analyst

Jack.li.2000@hotmail.com | (437) 265-0290 | <https://www.linkedin.com/in/jack-li-2000>
<https://Jack-Li-2000.github.io/JackLiPortfolio/>

Skills

Python: Highly proficient with pandas, numpy, matplotlib, seaborn, scikit, scipy, etc.

R: Highly proficient with ggplot, glm, scripting, visualizations.

Excel: Proficient with pivot tables, formula, vlookups, xlookup, conditional formatting, charts, data cleaning, etc.

SQL: Experienced with query calls, views, joins, conditional statements, database management, etc.

Data Visualization: Familiar with telling stories using interactive dashboards in Tableau and PowerBI.

Statistical Methods: Highly proficient with regression, classification, time series, hypothesis testing, and ML.

Cloud computing: Experienced with Google Cloud Platform (GCP) and Azure.

Work History

Data Science Intern, Teck Resources

(2021 – 2022, Vancouver BC)

Machine Failure Classification

- Performed data cleaning, conveyed statistical inferences and exploratory analysis with python and SQL.
- Retrieved relevant features for classification using forward stepwise, lasso, and elasticnet feature selection.
- Identified equipment failure using classification models validated with F1-score of 0.85.

Engine Life Cycle Prediction

- Predicted equipment lifecycle, finding trend, best, worst case with regression models.
- Validated projected engine lifecycle model prediction results with acceptable RMSE and MAPE
- Displayed insights with analytics dashboards and visualizations in PowerBI.

Projects

Toronto Real Estate Regression Analysis & Prediction

Using available independent data such as list price, # of rooms, etc., and target sold price for regression model to describe relationships and predict price, evaluated with RMSE in Python.

Wordnet Similarity Recognition Analysis

Using available data for human reaction time and accuracy based on response to stimuli, analyze correlation between Wordnet path similarity and human task performance with statistical testing in Python.

Toronto AirBNB Popularity Insights Dashboard

Interactive Tableau dashboard providing insights for features like popular neighbourhoods, pricing, accommodation types, etc. using data for the greater Toronto area between November 2022 and November 2023.

Global COVID-19 EDA & Visualization

MS SQL database manipulation and analysis of global and localized infection, mortality, vaccination, and vaccination survival rates by country, continent, and worldwide. Data imported, visualized as Tableau dashboard.

Interactive Bike Sales Analysis & Dashboard in NA and Europe

Excel data cleaning, analytics, and visualization of features contributing to bike sales in households for North America and Europe.

Education

Honours B.Sc. in Statistics & Computational Cognitive Science

(University of Toronto, Toronto Ontario)

Relevant skills include practical knowledge in data regression models, predictive data techniques, classifiers, time series, dimensionality reduction, data visualization, and drawing key information to deliver insights on complex data sets. Also, in-depth knowledge of data structures and programming.