

PETE PHUAPHAN

Toronto, ON, Canada | **Mobile:** +1(647)913-5248

Email: pete.phua@outlook.com | **Website:** petephua.com | **LinkedIn:** LinkedIn.com/in/petephua

EDUCATION

Humber Institute of Technology and Advanced Learning, Toronto May 2023 - Dec 2024(expected)

Ontario Graduate Certificate, Business Insights and Analytics

- Relevant coursework: Python, SQL, Data warehousing, ETL, Apache Spark, Hadoop, Machine Learning, Predictive Analytics, and Quantitative Statistical Methods

HarvardX, edX

Apr 2024

Data Science Professional Certificate, Machine Learning

- Relevant coursework: Machine Learning Basics, Linear Regression for Prediction, Cross-validation, k-Nearest Neighbors, Model Fitting, and Recommendation Systems

CERTIFICATION

- **AWS Cloud Practitioner Certification** (Validation Number: 5762871b31764d64af630cb505c8f48b) - Accredited for foundational AWS cloud knowledge and skills.

PROJECTS

Quantum Transaction and Customer Data Analysis Using R

Feb 2024

Analyze transaction and customer data **to identify** trends and inconsistencies.

- Developed R script to wrangle and analyze 264,837 transactional data entries to gain insights into the sales performance of 114 products across 21 target user groups.
- Identified that mainstream young singles and couples are 23% more likely to buy Tyrrells chips than other groups, leading to a strategic recommendation to boost overall sales performance.

Shopify Data Engineer Intern SQL Assessment

Jan 2024

Develop SQL queries **to uncover** insights for diverse analytical tasks.

- Utilized advanced SQL techniques, including window functions, CTEs, and date-time manipulation, to extract sales rankings, analyze purchase patterns, and identify the highest-grossing categories, potentially guiding the development of tailored marketing campaigns.
- Managed SQL queries within a GitHub repository for effective version control.

Canada Inflation Indicators ETL Using dbt

Dec 2023

Utilized dbt **to streamline** Canada's inflation data analysis.

- Initiated a complex ETL data pipeline project using dbt for systematic data transformations, targeting Canada's economic indicators including CPI, unemployment rates, and interest rates.
- Implemented structured transformations and conducted data quality testing across 4 key datasets within Google BigQuery using dbt, employing Jinja templates to enhance SQL code efficiency.
- Resulted in the successful transformation and integration of over 100 monthly data points for comprehensive economic analysis.

Iowa Liquor Retail Sales Seasonality Using Fourier Transforms

Dec 2023

Utilized Fourier transforms **to capture** sales seasonality.

- Leveraged Google BigQuery via an API from a Python script to query a comprehensive dataset containing 28 million transactional records from over 2,000 products.
- Developed a Python script to determine the optimal period and order values for Fourier transformations used in DeterministicProcess to capture seasonality in sales data from 3,063 stores.
- Achieved an RMSLE value of 0.0997 after fitting a Linear Regression model.

Kaggle Python Machine Learning

Oct 2023

Train a Python machine learning model **to predict** whether a passenger has been transported.

- Conducted Exploratory Data Analysis (EDA) and handled missing values using functions from Python's pandas library. Encoded categorical data with OneHotEncoder and applied StandardScaler to numerical data. Trained a supervised machine learning model using CatBoostClassifier.
- Achieved accuracy score of 0.8073, which is in the 91.8th percentile in the competition.

EXTRACURRICULAR ACTIVITIES AND INTERESTS

Exploring and expanding knowledge of modern data stack, including Databricks, Snowflake, dbt, Airflow, Terraform, and data lakehouse architectures. Also, a Manchester United fan and proficient in Thai cooking.