

# KAWALPREET DEOL

 LinkedIn |  Portfolio Website |  GitHub

## Skills

- Python | JavaScript | React | SQL | PySpark | SparkSQL | NoSQL | MySQL | SSMS | Tableau | Unix | Git | Jira | Confluence
- Azure (Data Factory, Databricks, Data Lake, Synapse, SQL Database) | GCP (BigQuery, Cloud Functions, Datastore, Firebase) | AWS (S3)
- Microservices | REST API | Frontend | Backend | Full-Stack | OOP | Data Modelling | Data Analytics | ETL | Data Lake | Data Warehouse

## Education

**Bachelor of Applied Science** University of Waterloo Waterloo, ON, CAN 09/2017 - 06/2023

- Major in Computer Engineering, Honours

## Experience

**Data Engineer, Co-op** Tealbook Toronto, ON, CAN 09/2022 - 12/2022

- Led a front-end solution enabling the upload of large data files to Google Cloud Storage and monitoring its data quality validation progress with status updates stored on Google Datastore using React.js and Python.
- Planned and conducted educational sessions, providing essential insights into the front-end, with a primary emphasis on React.
- Debugged microservice bugs by logging data objects on Google Cloud Logging.
- Streamlined Google Sheet reviews by developing a microservice to cleanse sheet data and append results to an aggregated sheet using Python and Flask.
- Triggered the microservice upon new file uploads with a Google Drive Trigger using Google Apps Scripts.

**Data Engineer, Co-op** Tealbook Toronto, ON, CAN 01/2022 - 04/2022

- Enhanced web crawlers to scrape online certification databases using Python, leveraging Scrapy and BeautifulSoup.
- Developed a script to extract JSON files from Google Cloud Storage, validating each file and logging the results using Python.
- Crafted a Python script using Pandas to cleanse JSON data, align it with a specific schema, store the resulting data frame in a spreadsheet, and send the sheet to an ML API via a JWT authenticated request.
- Employed Python to build a BERT model with data from Google BigQuery, predicting Small Business certification eligibility for companies with 82% accuracy.

**Full-Stack Developer, Co-op** PartnerStack Toronto, ON, CAN 05/2021 - 08/2021

- Contributed to PartnerStack's Design Sprint for the Partner Summary Page, involving goal definition, audience analysis, brainstorming, product prototyping, vendor validation, and design finalization for the first iteration.
- Utilized Test Driven Development (TDD) principles to reduce errors by building test cases in Jest that provided over 80% code coverage.

**Web Developer, Co-op** Horizn Studios Toronto, ON, CAN 09/2020 - 12/2020

- Created site routes, relational models, migrations, seeders, and controllers to build a front-end to show user topics using React.js, Laravel, MySQL, and S3.

**PHP Developer, Co-op** Venngo Inc Toronto, ON, CAN 01/2020 - 04/2020

- Developed a module enabling vendors to visualize product data statistics, access account details, and submit forms using Vue.js, Laravel, MySQL and S3.

**IT Technical Developer, Co-op** City of Toronto Toronto, ON, CAN 05/2019 - 08/2019

- Engineered a MySQL backend using relational modelling to manage the statuses and versioning of engineering design documents.

## Projects

- **Personal Website:** Crafting a personalized portfolio website to showcase areas of interest, projects and an experience timeline utilizing React.js, Material UI, and Firebase (Hosting, Functions, Firestore) (11/2023 – Current)
- **Banking Client Analysis:** Performing data cleansing, feature engineering and exploratory data analysis using Python in a Jupyter Notebook, leveraging Pandas, Scikit-Learn, Seaborn, and Matplotlib (12/2023 – Current)
- **IMDB Analysis:** Analysis of IMDB dataset to derive insights on the IMDB score and revenue using Tableau (11/2023)
- **Vehicle Telemetry Processing:** Created an ETL pipeline to process vehicle telemetry data retrieved from S3 and store the results in a SQL Database using Data Lake Gen2, Data Factory, and Functions App (10/2023)
- **Product Data Validation:** Built an ETL pipeline to validate product data using Databricks, Data Lake Gen2, and Data Factory (10/2023)
- **Pong Game:** Developed Pong utilizing Object-Oriented Programming (OOP) in Python, leveraging the Turtle Graphics library (08/2023)

## Certifications

- **Microsoft Azure Data Engineer Associate:** Evaluates expertise in data pipelining, storage, and streaming on Azure (EXP: 14/08/2024)