| **Junior Data Engineer Programme - Final Project** | |
| --- | --- |
| **Project Description** | Olist Store is the largest department store in Brazilian marketplaces. Olist connects small businesses from all over Brazil to channels without hassle and with a single contract. The Brazilian ecommerce public dataset of orders (from 2016 to 2018) made at Olist Store is provided to your company for analysis.  Your manager is asking you to critically analyse the provided datasets using Business Intelligence tools and provide some marketing findings / recommendations in a report format. The dataset has information of 100k orders made at multiple marketplaces in Brazil. Its features allow viewing an order from multiple dimensions: from order status, price, payment and freight performance to customer location, product attributes and finally reviews written by customers. A geolocation dataset that relates Brazilian zip codes to lat/lng coordinates is also integrated in the dataset.  After a customer purchases the product from Olist Store, a seller gets notified to fulfill that order. Once the customer receives the product, or the estimated delivery date is due, the customer gets a satisfaction survey by email where they can give a note for the purchase experience and write down some comments.  The dataset is available [here](https://drive.google.com/file/d/1bLwp3KmwvQHB2ucquErlkayI8yCEvmO9/view?usp=sharing) (please download the ZIP file)  **1. Data Schema**  The data is divided into multiple datasets for better understanding and organization. Please refer to the following data schema when working with it:    **2. Attention**   1. An order might have multiple items. 2. Each item might be fulfilled by a distinct seller.   **3. Possible scopes of the marketing findings / recommendations from the dataset:**   1. Feedback Sentiment Analysis: Evaluate the polarity of the tweets as customer feedback positive, negative or neutral. 2. Clustering: Some customers did not write a review. But why are they happy or mad? 3. Sales Prediction: With purchase date information you will be able to predict future sales. 4. Delivery Performance: You will also be able to work through delivery performance and find ways to optimize delivery times.   **4. Reference**   * Sentiment Analysis in Power BI:<https://www.youtube.com/watch?v=mhe9Vs3jQes> * Text Mining and Sentiment Analysis: Power BI Visualizations:<https://www.red-gate.com/simple-talk/sql/bi/text-mining-and-sentiment-analysis-power-bi-visualizations/> * Forecasting in Power BI:<https://www.youtube.com/watch?v=mMd2rKK7dWc> * Power BI Clustering:<https://www.youtube.com/watch?v=LsFDWS77ER4> |
| **Project Objectives** | 1. Create data pipeline to ingest data in PostgreSQL or SQL server database 2. Use Power BI to answer following questions:    1. How many customers, orders, and orders per customer does the company have?    2. What is the number of customers by state?    3. What is the number of orders by month?    4. What are the top 5 product categories? 3. Visualise the company’s customers’ demographics, sales trend, orders by categories, orders changes by year, etc. and use Power BI to help make better decisions 4. Map and compare report data with data from database query to validate the reports (functional testing). 5. Critically analyse relevant data using statistical methods (e.g., Predictive Modelling or Machine Learning) 6. Provide some recommendations and improvements (please refer to point 3 in the project description) |
| **BSM** | All (see [BSM Evaluation Rubric](https://docs.google.com/document/d/1TFShS188dINbDR3SZU2KZBdqHnfJIJ_9Ylxe0CleGH8/edit?usp=sharing)) |
| **Duration** | * Final Project: 15 hours + Presentation: 20 - 30 minutes per team |
| **Project Team Structure** | * Small groups of 3 - 4 members |
| **Project Structure** | The project will be organised in **3 different phases**   | **Project Phase** | **Focus** | | **Session Duration (hours)** | | --- | --- | --- | --- | | **Phase 1** | Project Briefing and Team Formation | | 0.5 | | Brainstorming | | 0.5 | | Analyse the dataset and data schema, import the dataset into the PostgreSQL, and set up the corresponding data schema | | 1 | | **Phase 2** | System Design   * Create and manage data relationships in Power BI Desktop * Design dashboard (various visuals, slicers and tabs) | | 1  8 | | **Phase 3** | Publish the dashboard to Power BI web service (optional) | | 0.5 | | Functional Testing & User Acceptance Test | | 1 | | Presentation Prep - Creation of presentation deck and rehearsal | | 2.5 | |
| **Project Deliverables** | 1. Your **project portfolio** should include the following:  | **Phase 1** | **Phase 2** | **Phase 3** | | --- | --- | --- | | * A list of ideas (e.g., what kind of visuals) generated, consisting of conceptual sketches, function and key components | * A document about the dataset, data relationship, and Power BI dashboard description * The Power BI Desktop source file or the published Power BI dashboard URL | * A user guide about how the dashboard works, e.g., what’s the meaning & relations of visuals, slicers, etc. * Dashboard |  1. **Presentation slides**    * Guide: [Guidelines to Create a Presentation Deck](https://docs.google.com/document/d/1UwFH_8KzTti7XCvyw06qfLWhdznMgnLuRRvxnTFUsBM/edit?usp=sharing) |
| **Presentation** | * Time is allocated for presentation prep and rehearsal before the actual presentation.   + Guide: [Tips to Deliver a Presentation](https://docs.google.com/document/d/1pWhRsFhhHjx66AJix9wTh8ZNgXFJHcjWNg7nPrUtfBU/edit?usp=sharing) * Each team will take turns to present the completed project to an assessment panel. * Final team presentation will take place in the final week of the program. * Potential employers may be invited on the presentation day. |
| **Assessment** | * **Formative Assessment of Project Deliverables** -   + Specific qualitative feedback will be provided by instructor(s) for each team’s performance (technical)/ deliverables throughout all phases.     - Download to fill up and submit this [Deliverables Checklist](https://docs.google.com/document/d/1aNftpkE3bqdCdgQ3xOKh7w4xiX3VnmCsmi545LvlU-0/edit?usp=sharing) at the end of each project phrase:   + Team score will be given for presentation based on a rubric.     - [Presentation evaluation rubric](https://docs.google.com/spreadsheets/d/19T_OLkFk1ifiI8nnkhsfOCal2S___-gySG1eutGmUSo/edit?usp=sharing) * **Peer Assessment** - Participants will assess their team members’ BSM and collaboration in the project.   + Download to fill up and submit this [Peer Assessment BSM Form](https://docs.google.com/document/d/1b4ROXTmMrneqKI9YyVlM_hPdo2ynH_5dlguqK8lYxJI/edit?usp=sharing) by the end of the project. * **Self-Reflection** - Participants will complete a self-reflection form to self-assess how they have demonstrated BSM in their project teams.   + Download to fill up and submit this [Self-Assessment BSM Form](https://docs.google.com/document/d/1c0Qltk2ljc8jd3X6cF9FAODEbjX0R4o7IKO-Z3NEGAw/edit?usp=sharing) by the end of the project. |