

**Assessment  
Foundation Data Engineering**

**Course Documentation**

**Student ID:** 2408643

**Student Name:** Sameul Bajracharya

**Section:** L6CG7

**Tutor:** Mr. Gunjan Kumar Mishra

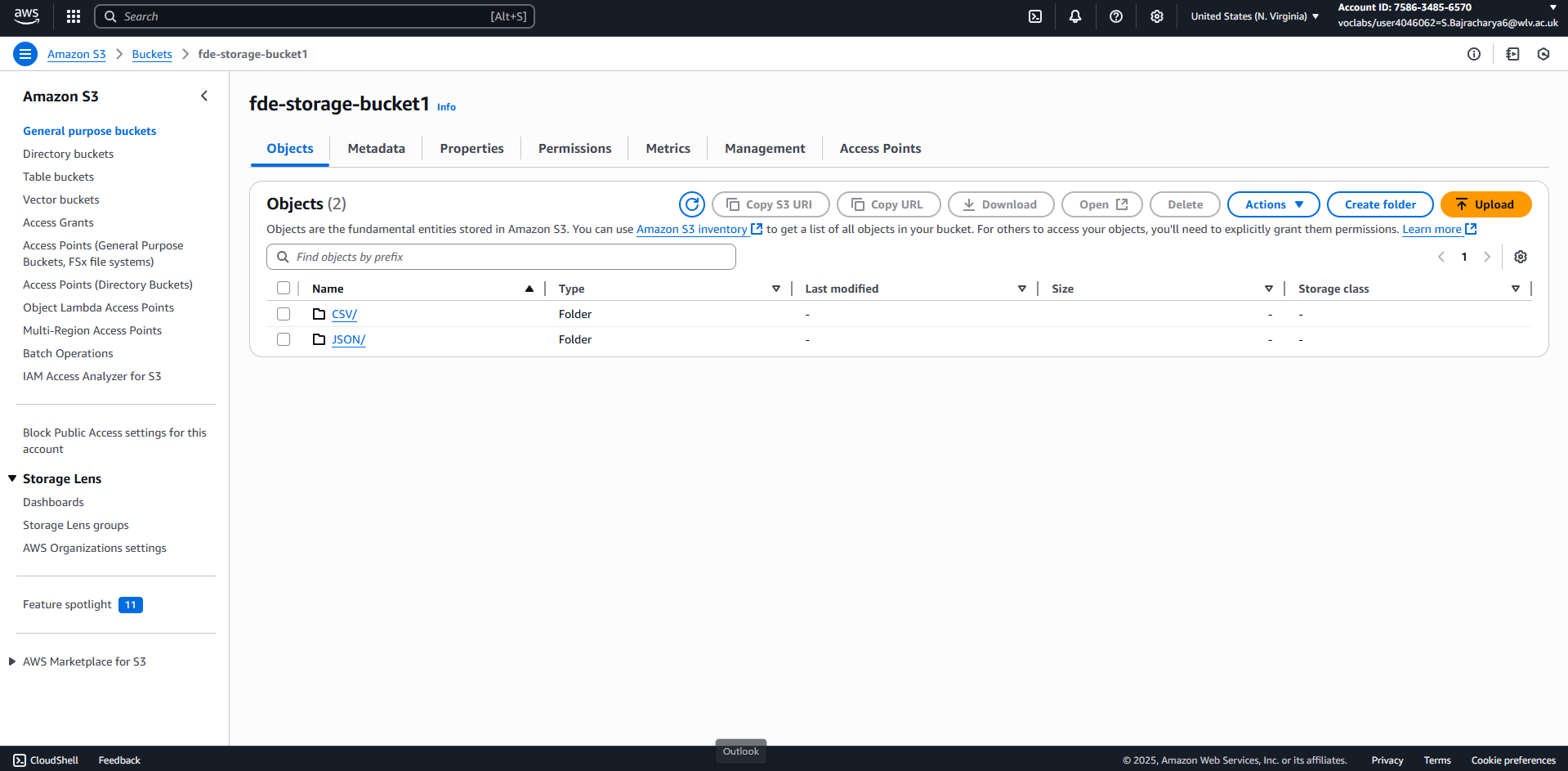
**Word Count:** 1300

**Submitted on:** 2025-08-25

# Week1

During the first week of foundational data engineering class, we learned about the tools and goals of our overall course for 6 weeks. We were introduced to the databases and cloud storage that we will be learning in the next few weeks. We also understood the prerequisites which are to be fulfilled for this course. Then, the tutor went on to give us an overview of data engineering. We understood that it includes working with data which is beneficial for analysis tasks. Then, we worked with AWS S3 bucket by first starting the learning lab and then creating a bucket. We practiced folders’ organization by creating two folders, one for CSV files and one for JSON files. We uploaded 3 files inside the CSV folder and 2 files inside the JSON folder. This way we learned about managing data in cloud storage and its importance.

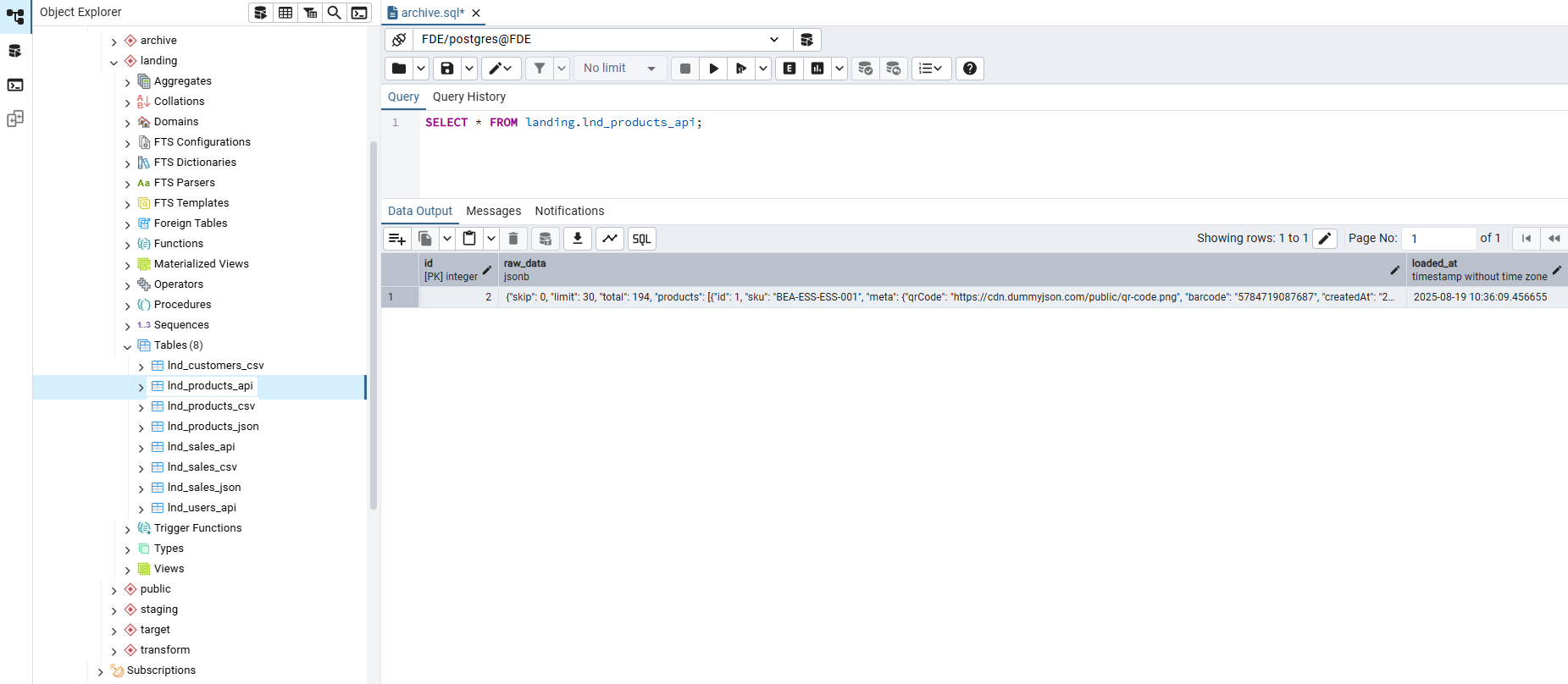
Continuing our first week, we went on to learn about relational databases and some queries. For that we installed PostgreSQL and pgAdmin to work with databases. We created a database first and then a schema inside it to make the concept clearer. Overall, throughout this first week, we learned about foundational data engineering , the mechanism of cloud storage and relational databases.



# Week2

During the second week of foundational data engineering class, we focused on creating landing tables and extracting data which needed to be filled into those tables. First of all, our tutor gave us a brief explanation about the landing tables and DDLs (Data Definition Language) which we would use to manipulate tables and create schemas in databases. We learned different basic DDLs and proceeded for hands-on practice. We wrote DDLs in pgAdmin to create schema for Landing and landing tables for understanding data organization in Landing stage.

We created a FDE folder and another folder named Extractor inside it where we kept all the required configuration files. In config.yaml, we configured all the necessities as provided to us by the tutor. In the next class of our second week, we now needed to extract the data stored in the AWS bucket to our landing tables using python. We updated the files with their respective codes and placed a requirements.txt inside that folder which consisted of all the necessary packages. We installed all the packages and ran all the files. Now, in our tables we were able to view data. This week helped us understand the combination of coding and databases in data engineering.



# Week3

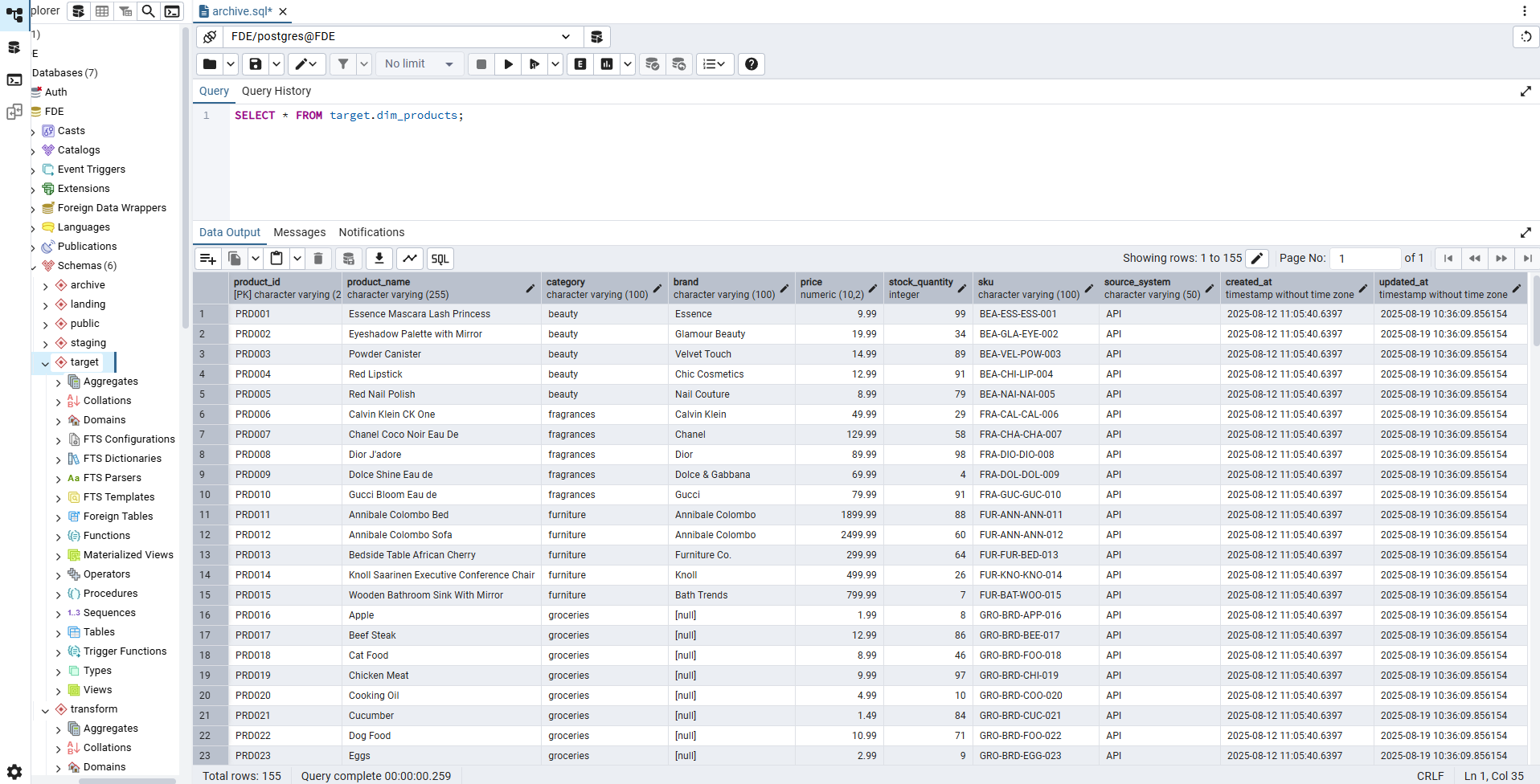
During week 3 of our foundational data engineering class, we learned about various advanced concepts present in SQL. This was done by creating a database named as advancedsql. This week helped us explore and learn about complex and advanced SQL not only basics. We learned about different topics throughout this week. We started off with a topic on schema separation where we created schemas for different purposes and also used DEFAULT keyword. Then we learned about preventing duplicated using constraints which enabled us to prevent the insertion of duplicate values which is essential if we work with larger datasets in the future.

Another essential concept we learned is unifying data from different sources in which two or more tables were combined. We learned about how to handle JSON data by creating table then inserting JSON data to it. Like this, we learned how semi-structured data can be stored using SQL. In advanced string processing, we worked with textual data and used different functions to work on it. After this, we focused on window functions which provided us with ways to calculate different functions in rows and preventing them from collapsing. Additionally, we learned about computed fields such as ANY() function to create dynamic values and also about complex data types. In next class, we were provided with sql queries which we ran to create different schemas(staging, transform and target) and tables inside them.



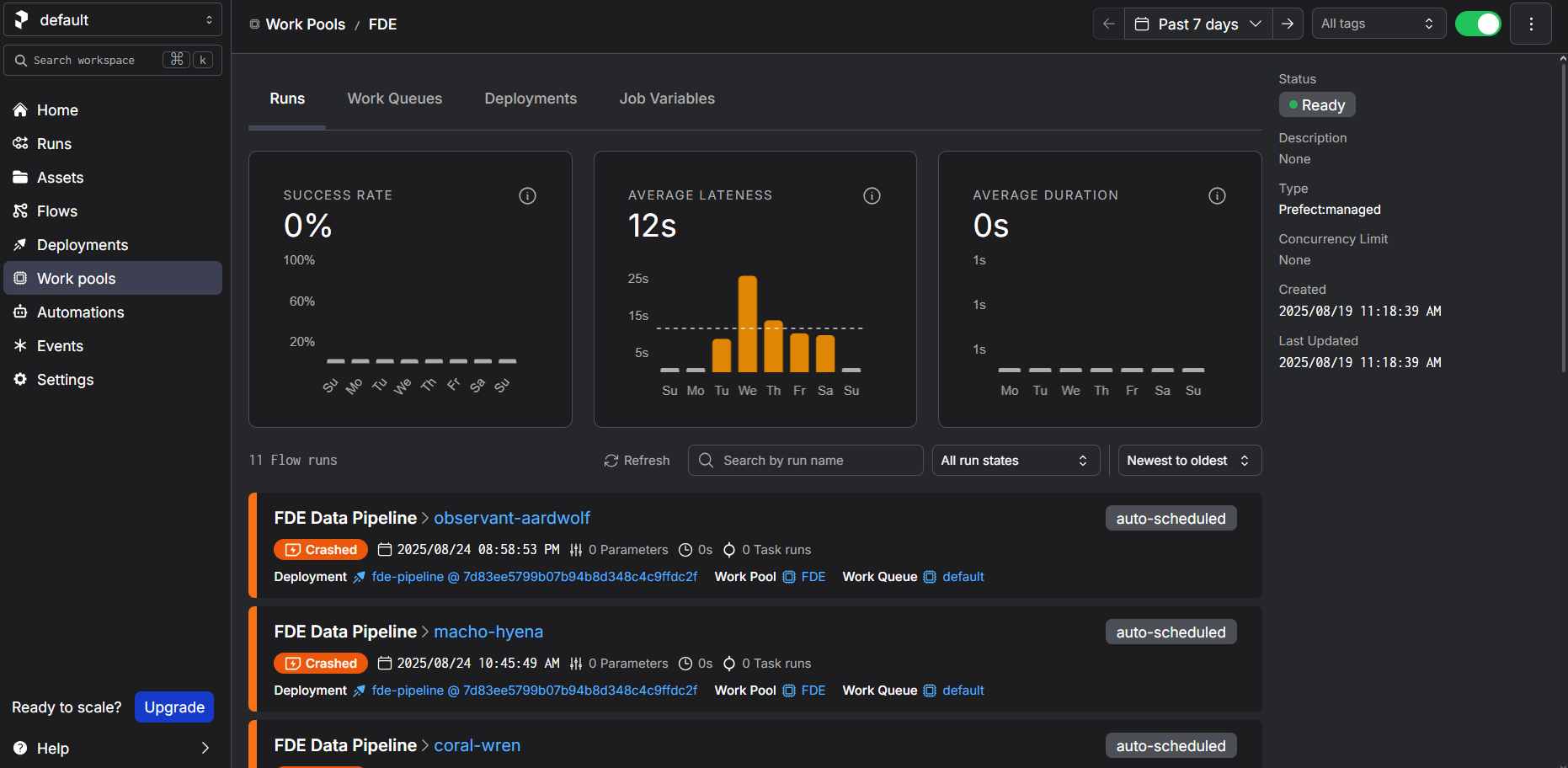
# Week 4

During week 4 of our foundational data engineering class, we continued our work on Advances SQL part 2 which helped us to explore different practical implementations for database management. We were introduced to upsert also known as the combination of update and insert. This is used for handling the records that may already exist in the tables. To learn the skill of data migration, we learned about loading data from one table to another. We continued our learning with deduplication which makes sure that the datasets remain clean and we used conditional queries to update the records only if some criteria are met. We learned about data enrichment which enhances data with more information. And we created temporary tables essential for data processing to store results before finally processing. In the next class of the same week, we focused on practical implementation of the knowledge we had acquired. We were provided with Loader code which we downloaded and placed inside the FDE folder. We then created a new virtual environment and installed the required packages. We learned about a concept known as ETL which is an essential topic in the field of data engineering. Finally, using the virtual environment, we ran the Python scripts. We observed the process of ETL pipelines and the movement of data across different systems.



# Week 5

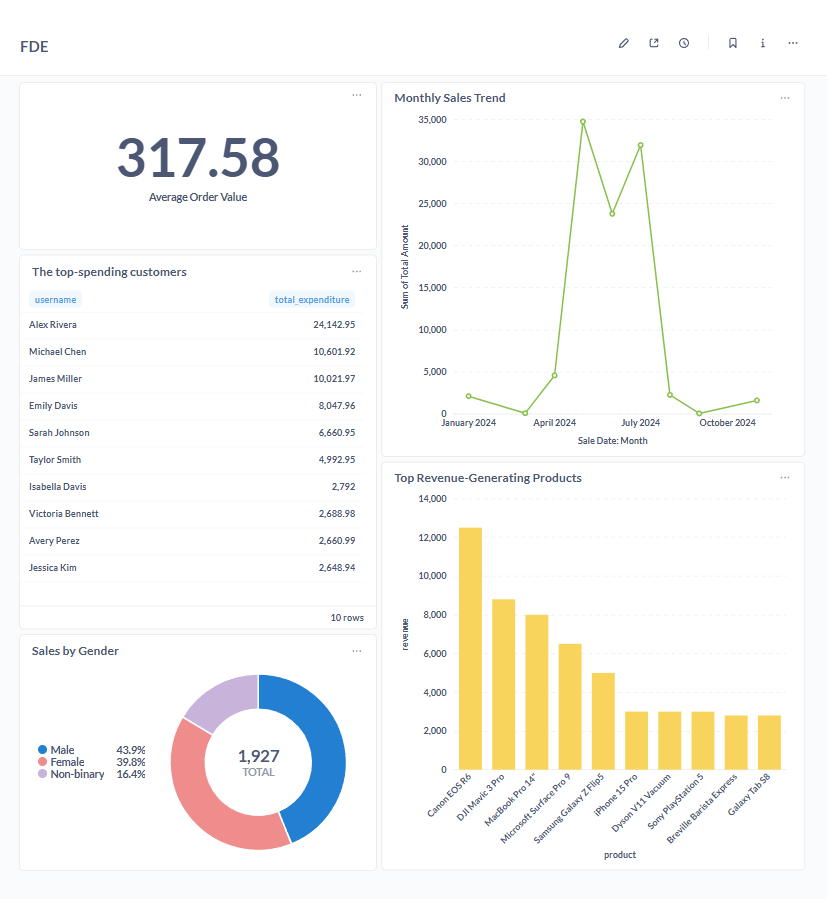
During week 5 of our foundational data engineering class, we learned about archiving and orchestration in data engineering. For managing large databases, archiving is a very important concept and process as it separates the data into different tables. It allows older data to be separated. This often helps the database to have a cleaner and consistent look. In first class of this week, we were introduced to archiving of data. For that, we downloaded the python script and sql file provided to us in our tutorial file. For the sql file, we ran the query in pgAdmin and created tables whereas we placed python code in our extractor and ran the script in our terminal. After that we viewed data inside all the archive tables to ensure that the data has been loaded successfully and archived correctly. In our second class, we were introduced to orchestration. A modern workflow tool known as Prefect was used as a tool for orchestration. In our workshop file, we were provided with pipeline.py to be placed inside the extractor. As provided in the documentation, we followed all the steps and ran the file using prefect. We observed the dashboard link provided in the powershell. This process was done locally. Now, we had to do this process online. For that, we created account in prefect cloud.



# Week 6

During our final week of foundational data engineering, we learned about two important aspects of data engineering. They were Docker and data visualization. Our tutor gave us a brief explanation of Docker and why it is used. We learned that Docker is used to create isolated environments which allows consistency in systems. For this, we setup Docker on our windows and after the installation process, we activated the fde virtual environment. In that environment, we installed metabase for visualization process.

Metabase is known as a powerful tool which is used to explore and visualize raw data into meaningful information. Once the metabase was installed, we went on web and opened the localhost for metabase. There we explored the home page and then connected the metabase to our database and started exploring to analyze data. Two of the business questions were already answered, one using a line chart and the other one using SQL query. We placed both the data in our dashboard and saved it. We proceeded to answer 3 more questions using aggregation average to find average value. We also create bar graphs and pie chart using SQL query. After the completion of all the questions, we managed the layout of our dashboard and saved it.

  
Link to github repo:

<https://github.com/SamuelBajracharya/Foundation-Data-Engineering-Course>