Prior Intro into Data Analyzation Segment:

1. Data Used: Our dataset from Kaggle.com is a simulated credit card transaction dataset covering the credit cards of 1000 customers from a pool of 800 merchants. The data set covers transactions over a 2 year period covering legitimate as well as fraudulent transactions.
2. Link to Dataset: <https://www.kaggle.com/datasets/kartik2112/fraud-detection>

Start Here ======================================================================

1. Data Review: The Kaggle data was reviewed and analyzed by our Machine Learning team and placed into 2 separate tables:
2. a. Model\_Upload table: Contains those attributes which can be used in machine learning models (slide 1)
3. b. Personal Info table: Contains those attributes which are more personal in nature (slide 2)
4. and is commonly referred to as “P.I”
5. A database was created in Postgres (named Credit\_Card\_Fraud) with the 2 tables created with this database.
6. Tables were then created by means of “Create Table” queries within the Credit\_Card\_Fraud database within Postgres.
7. A database schema was created which displayed the tables, attributes and appropriate data types and characteristics assigned
8. And a Entity Relationship Diagram was created (slide 3)
9. Data suiting the appropriate columns were created as CSV files and then uploaded within Postgres to the appropriate 2 tables
10. Data Testing: The data was then tested and reviewed by means of creating joins within the tables and related queries such as performing counts on specific types of data
11. Sample files: Sample files were created to be placed in GitHub as there is a size restriction on the file that can be placed in there: (Initial 554,000 rows of data cut to 100,000)

(model\_upload\_s)  
(personal\_info\_s)

1. Python code introduced where:

a. Username/Password of database in Postgres is checked  
b. Tables are read in from Postgres

1. Amazon Web Services: After the review steps outlined above, the 2 tables were then placed in the S3 area within AWS to be used in our machine learning analysis
2. Sean will now review our Machine Learning Analysis

**a**