

Presentation Outline

1. Introduction (1 mins) - Rabia
 - The topic and reason for its selection
 - Description of the source of the data
 - Questions we hope to answer with the data
2. ETL Dataset (1min) - Arthur
 - Describe nature of dataset
 - Process of extracting from an S3 bucket and putting it into a pandas dataframe
 - Transforming the dataset with sklearn label encoder
 - Dividing the dataset into male/female and buckets of age
 - Loading the dataset into the database with an SQLAlchemy connection string
3. Machine Learning Models and Analysis (2 min) - Asad
 - Four different machine learning models chosen
 - Process of running the data through the machine learning model
 - Results: Accuracy, Precision and Sensitivity
 - Importance of Features Analysis
 - Final model chosen
4. Database (1 min) - Arthur
 - Relational Database is hosted on AWS RDS and created on PostgreSQL
 - Describe entity relational diagram (ERD)
 - Static storage for individual tables
 - Dynamic storage for user inputs from Dashboard
5. Demonstration of Dashboard (2 mins) - Rabia
 - Spend a couple minutes opening up dashboard from Github repo link in ReadMe and going through the form, submit and explain results
6. Overall Project Architecture (2 mins) - Asad
 - Describe the user input data pipeline:
 - Extracted from HTML form and uploaded to S3 via AWS SDK for JavaScript in Browsers
 - AWS Lambda function one, extract from S3 bucket and upload to database
 - AWS Lambda function two, extract from database, pass through RandomForest model and return prediction and importance of

features to second S3 bucket

- Fetch the results and features JSON files in Dashboard script and display in HTML as naked JSON

7. Anything we would have done differently (1 min)

- Asad:
 - Never quite figured out a way to display the actual dataset in a useful manner on the dashboard due to it being perfectly encoded for input into ML models when we received it
- Arthur:
 - Would improve dashboard aesthetics with more Bootstrap and CSS styling
 - Add simple definitions for each symptom for ease of use
- Rabia:
 - Add graphs to display dataset onto dashboard, as well as ML model analysis.

7. Q & A (1 - 2 mins)

Total Time = 10 - 12 mins