Coding challenge

Azure Databricks

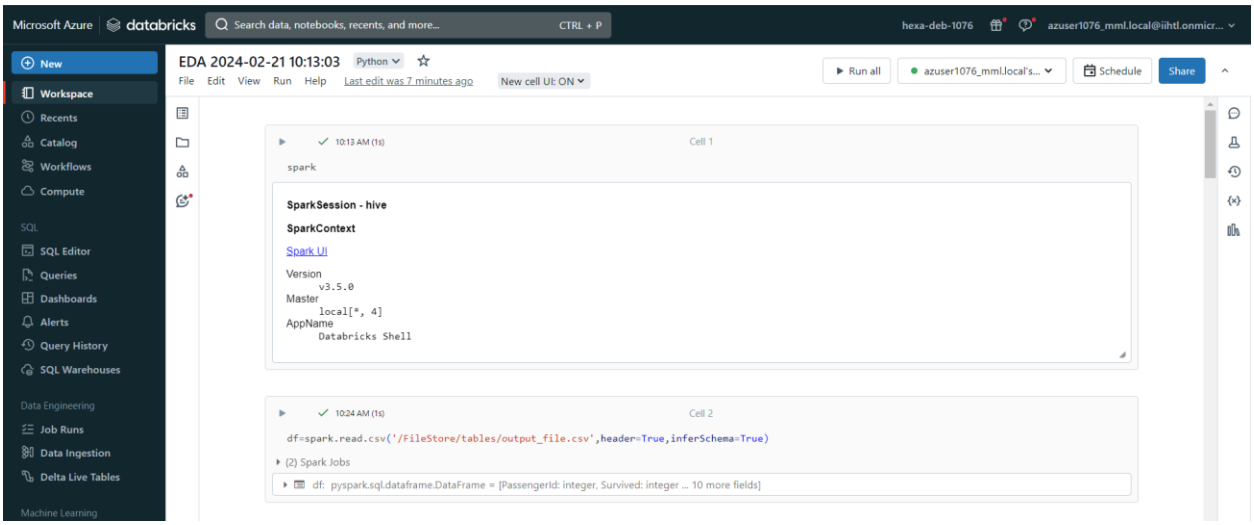
Name: Rohan Vinayak Chaudhari **Email:chaudharirohan24@gmail.com**

Batch: Data Engineering 1

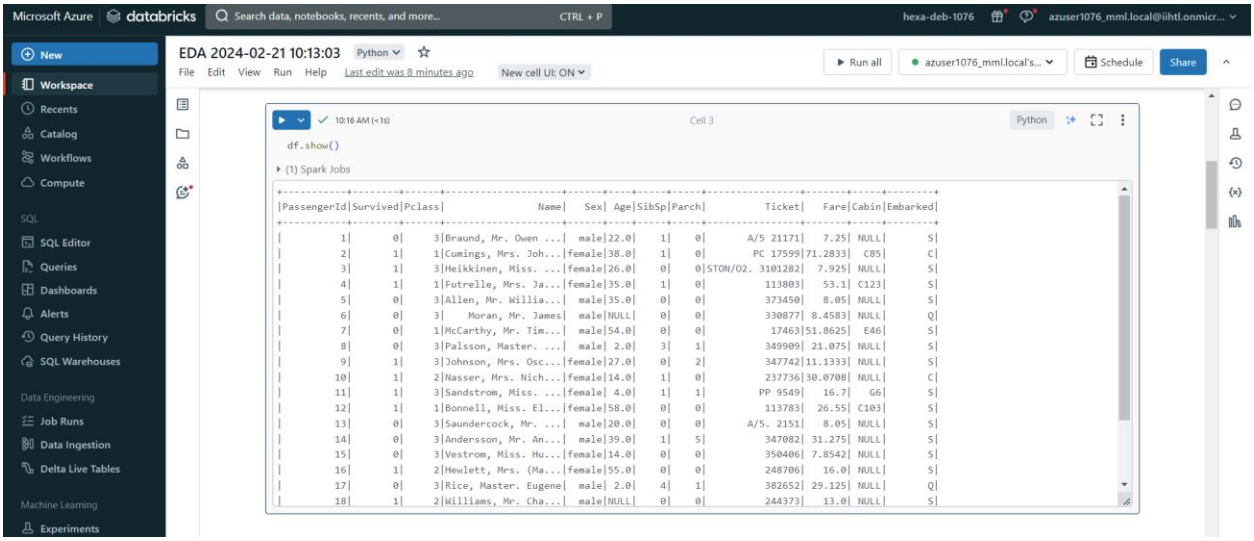
Question1: Exploratory data analysis (EDA) in Databricks &Visualizing data in Databricks.

Spark Session:

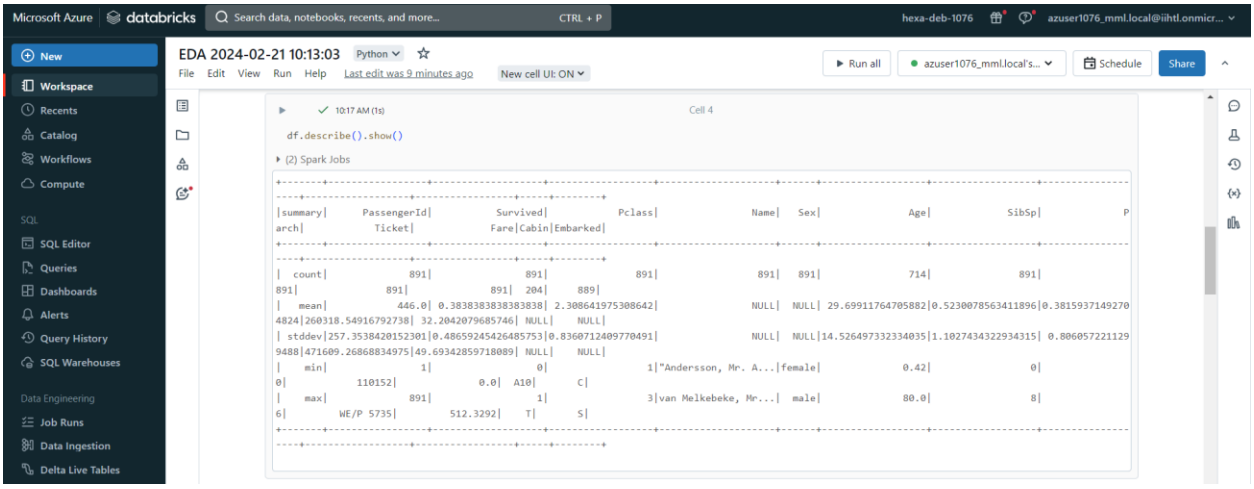
Uploaded a Titanic dataset for EDA:



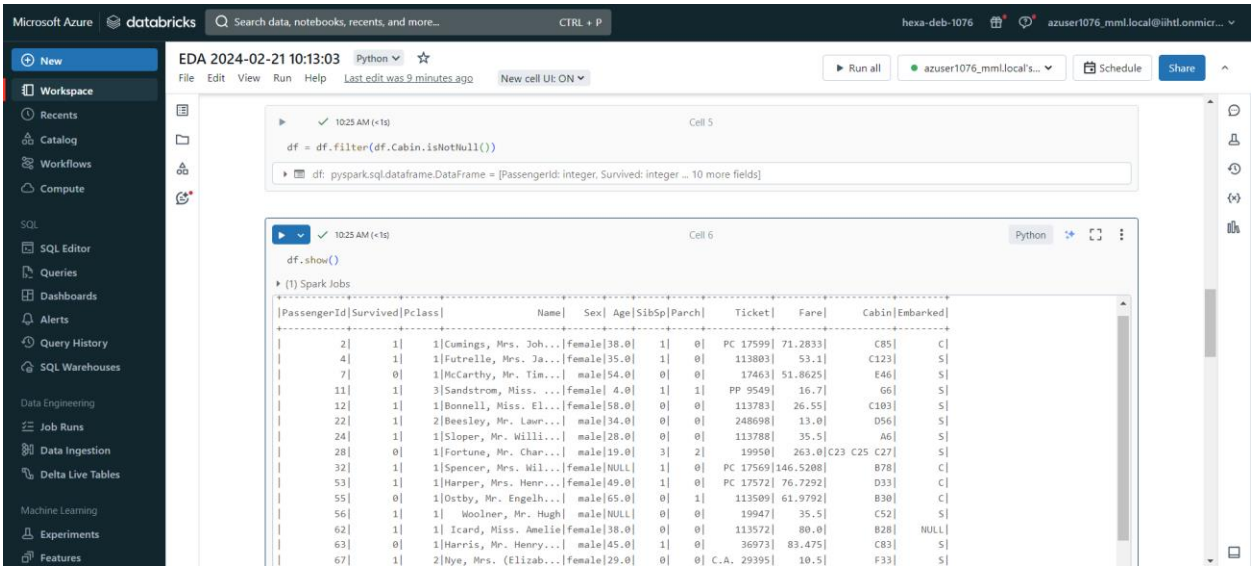
Viewing the data:



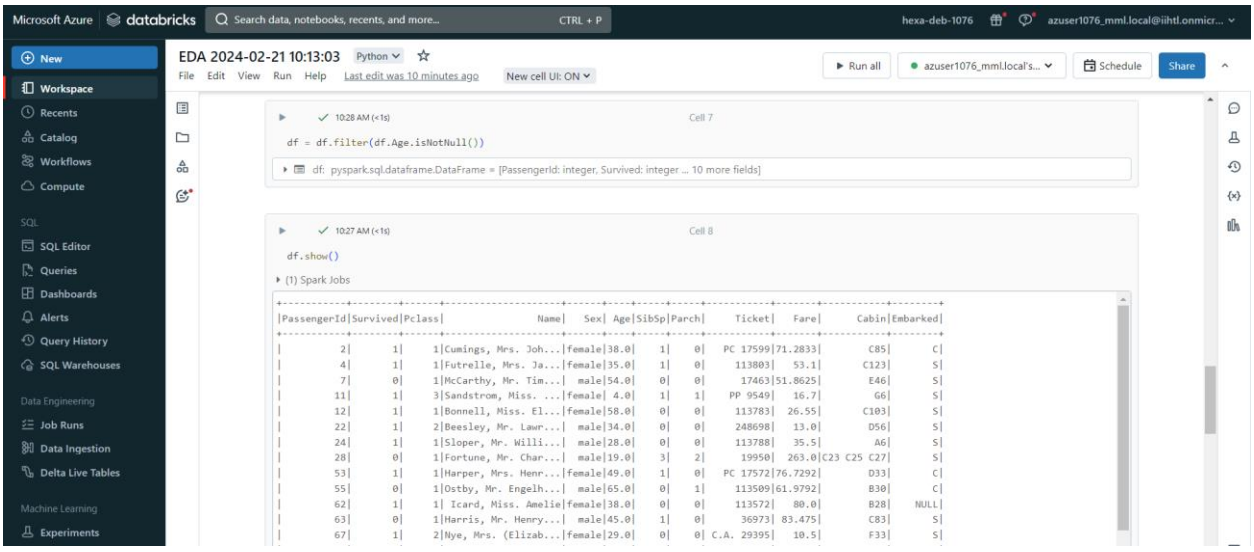
Describing the whole dataset:



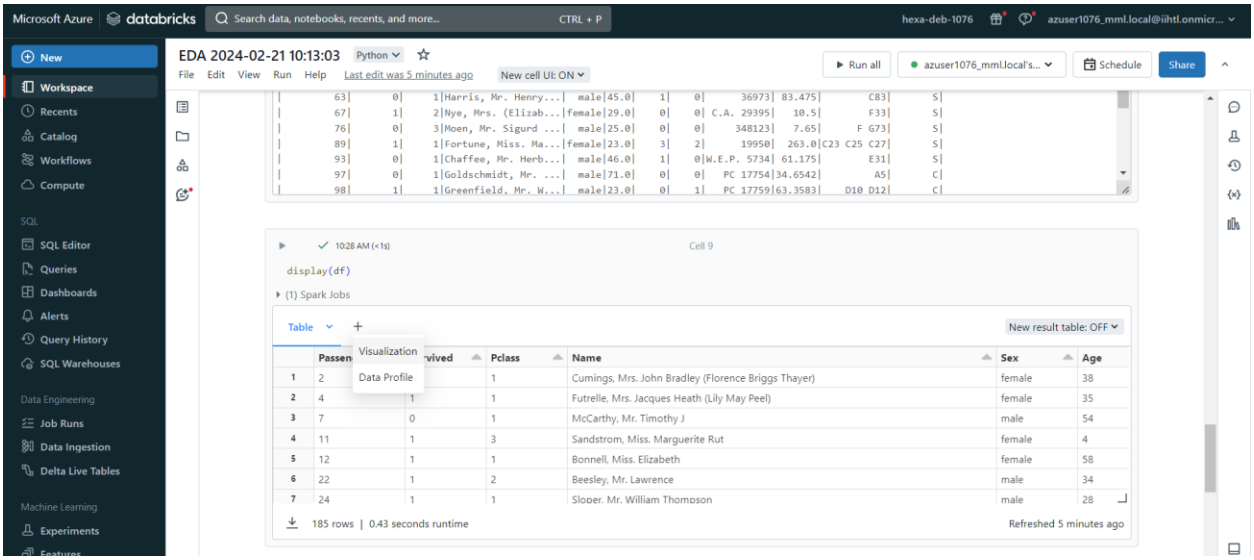
Filtering Cabin column which contain null values:



Filtering Age column which contain null values:



Using Display method for Analyzing the data:



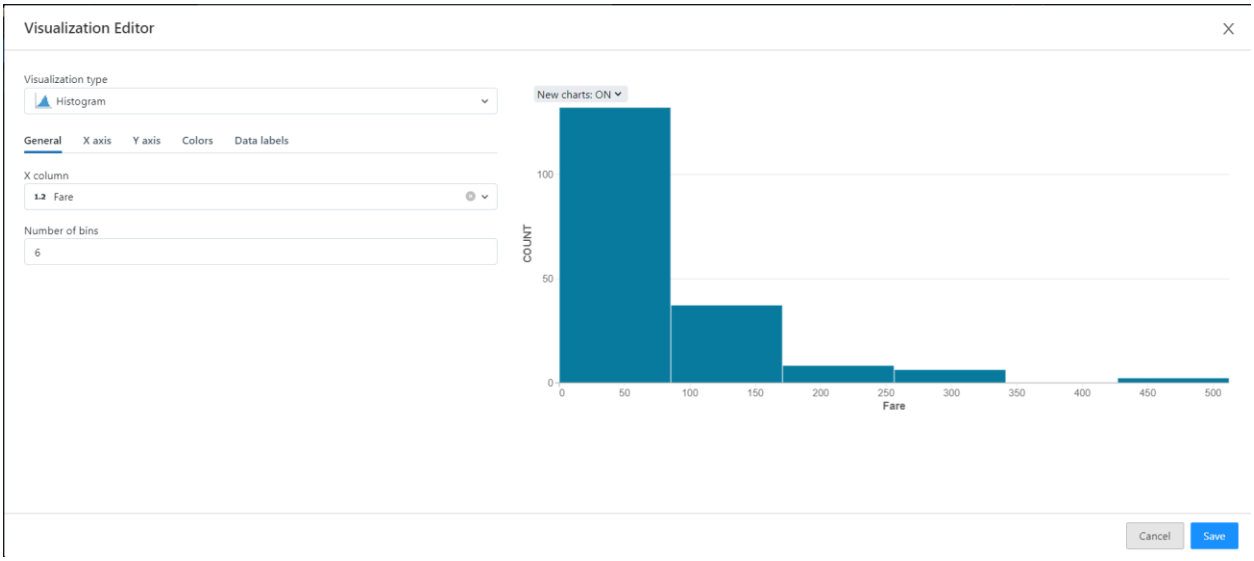
Analysis

Analyzing How many people survived of particular age group and grouped by Gender:

Bar Graph

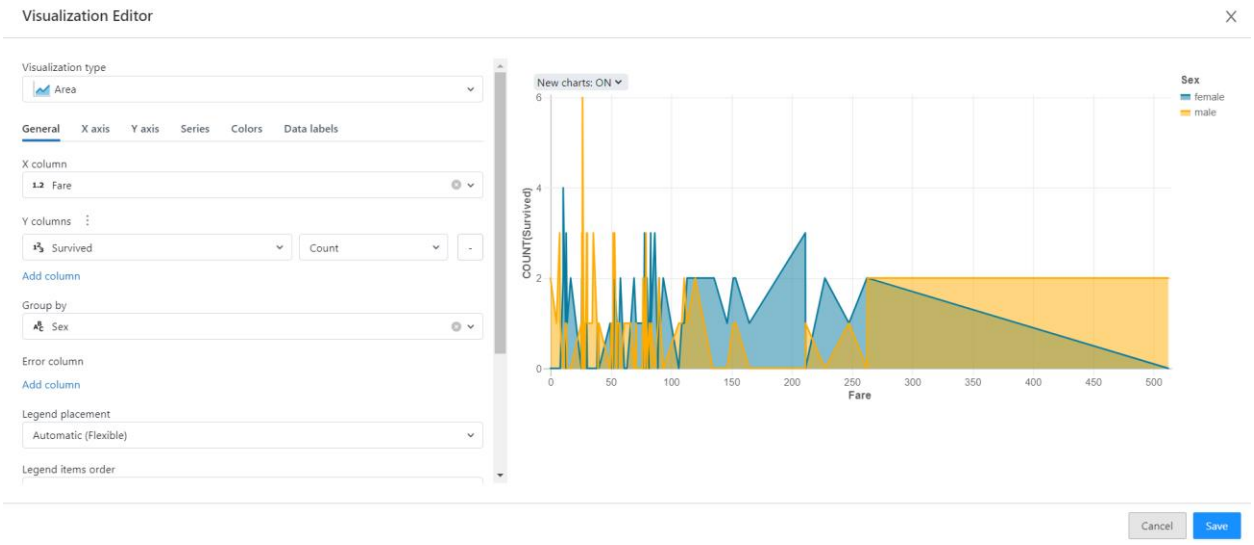
Count of people lying in particular Ticket price:

Histogram:



Analysis for <u>fare vs survived</u> where we will come to know that higher ticket price passenger survived more or not:

Area chart:



Here we will come to know about the proportion of the gender:

Pie chart: