Project Statement:

Task 1: Data Cleaning with Azure Databricks

Dataset: <https://raw.githubusercontent.com/justmarkham/DAT8/master/data/chipotle.tsv>

Tasks: Clean the dataset and perform the following using Databricks:

What is the number of observations in the dataset?

What is the number of columns in the dataset?

Print the name of all the columns.

How is the dataset indexed?

Which was the most-ordered item?

For the most-ordered item, how many items were ordered?

What was the most ordered item in the choice\_description column?

How many items were ordered in total?

Turn the item price into a float

Check the item price type

Create a lambda function and change the type of item price

Check the item price type

How much was the revenue for the period in the dataset?

How many orders were made in the period?

What is the average revenue amount per order?

How many different items are sold?

Task 2: Machine Learning with Azure Databricks

Create a Machine Learning model to find the diagnosis of Breast Cancer Patients based on the features in the dataset below.

Dataset: https://raw.githubusercontent.com/deepanshuMeteor/ComputerVision/refs/heads/main/data.csv