Q1 Data Science Phases

6 Points

As you must have learned from the videos, the process of data science involves multiple phases. In this assignment, imagine that you are a data scientist at a restaurant chain. Using the prompts below, provide your thoughts as you go through the phases of the data science process. The provided example walks you through one use case for all the phases. Use your creativity to come up with your own use cases for the restaurant chain. Answers should be brief.

Q1.1 Asking the right questions

1 Point

Before we start a data science pipeline, we learned in Module 4 that asking the right questions is helpful to extract actionable insights. For a restaurant chain, provide 2 potential questions for creating data products. For example, which localities should I market the brand of the restaurant chain?

Sample questions(but not limited to) :
1. What are some new recipes that the restaurant chain can come up with?
2. Where can we open more stores?
3. How can we increase revenue? - item costing, staffing costs etc.

Where is our most popular location?
What demographic is frequents our restaurant chain the most?

Q1.2 Acquire

1 Point

For the questions above mention briefly the kind of data that you would require. You can mention fields that would help you in the task. For example, customer demographics like age, locality of residence, occupation etc, other restaurants in the area would be useful data for finding which localities to market the restaurant in.

Where is our most popular location?: sales across all store locations
What demographic is frequents our restaurant chain the most?: customer demographics across all store locations

Q1.3 Prepare
1 Point

What kind of data preparation do you foresee for your questions. For example, some data transformations could be required to group some zip codes into a locality. If I do not have the age of a customer, I could use the average age of the occupation.

Where is our most popular location?: convert sales to one standard currency
What demographic is frequents our restaurant chain the most?: convert zip codes to strings instead of integers

Q1.4 Analyze
1 Point

Now that you have prepared the data, you can perform analysis to generate actionable insights for your questions. What could be some possible analysis steps for your questions? For example, I want to classify whether a customer would purchase at the restaurant, or, I want to know which category of customers buy at similar restaurants.

Where is our most popular location?: sum total sales for each location
What demographic is frequents our restaurant chain the most? get count for each zip code

Q1.5 Report
1 Point

What would be some useful ways to report the analysis outcomes for your questions? For example, correlation of data fields, histogram for popular localities, etc.

histogram for total sales across each location
histogram for total count for each zip code

Q1.6 Act
1 Point

Given the analysis and report from data science experts and the initial questions, as the restaurant chain owner what are the actions you can now take? For example, if the reports suggest that chain is popular among students, I can invest in marketing the brand on university campuses.

close stores that have the lowest sales and close stores where there is the least amount of foot traffic by zip code

Q2 Python Setup
4 Points

Please submit a screenshot of the Google Colab environment for your Week 1 Assignment, in which you are running a cell. It doesn't matter what cell you run.

Screenshot 1_12_2025 8_20_14 AM.png Download