AccelerateAI

Data Science Global Bootcamp

Python Data Analysis: Assignment 01

Q1: Consider the Credit card dataset provided. The file ***Credit Card Data.xlsx*** has 3 tabs with the following data:

- Credit Card customer master data
- Spend (Transaction) data
- Card repayment information

Read the data in the excel file using Pandas library and perform the following:

1) Check the data for any corrupt data, missing values, incorrect data types etc.
2) Make the following substitution:
   a. Age of credit card customer can't be less than 18. In case age is less than 18, replace it with mean of age.
3) Answer the following Questions:
   a. How many unique customers are there?
   b. How many spend categories are there?
   c. Which category has the highest average spend?
   d. What is the average monthly spent by product categories?
   e. Which customers are reaching 90% or more of their spending limit?
   f. Which city has the maximum number of spenders, each month? Is there a need to run campaigns in specific cities?
   g. Which age group spends the most?

Q2: The demand and usage of the internet is growing across the globe and public wifi is becoming a new normal as we move forward. Wifi providers want to optimize the cost while providing the free and paid wifi. The use case here refers to public WiFi hotspots in New York City. Please refer to the dataset – **Wifi Data.xlsx**

- The information comprises of records for every public WiFi hotspot (ones provided by or in partnership with the city) in New York City.

Using Python's Pandas and Seaborn library to answer following business questions.

1) Get the data provided, check frequency distribution of the features, drop irrelevant and redundant features if any and state your inferences?
2) Using appropriate data visualization answer the following:
   a. Plot the providers vs number of all wi-fi hotspots they have in operation? Who has the highest installations?
   b. Plot providers vs number of **free** wifi hotspots.
   c. Are there any provider(s) who do not provide free wifi?
   d. Which Location type (Location_T) are most common and least common?
   e. How many free "indoor" and "outdoor" hotspots are there? Which companies have the highest number of Outdoor Free installations.