Question 1:

Calculate the percentage of missing data in the given dataset based to variables

Answer

Origin Country 0.0
Origin Station 0.0
Origin Airport 0.0
Destination Country 0.0
Destination Station 0.0
Destination Airport 0.0
Customer Name 28.512277148953096
Customer Market Group 28.512277148953096
Chargeable weight 0.0
Actual weight 0.0
Cubic meters 0.0
Airline 0.0
Service 1.7484540556178354
Pickup 30.508443930134163
ReceivedAtStation 0.11391169131739776
ArrivedAtDestAirport 2.742921202039562
ArrivedAtDestStation 9.145481502911077
OutForDelivery 35.976205113369254
Delivery 4.6812280765197265
Transit Time Requirement 33.057896069142586

Question 2:

Split the data based on different Customer market groups

Answer

Data splitting is done using pandas library using loc function, the following syntax is data.loc[data['Customer Market Group']==' name of market group ']

please refer to the code for details and can generate by running the code.py file

Question 3:

List the names of customers in different customer market groups

Answer

This is done by using the unique function

please refer to the code for details and can generate by running the code.py file

Due to lack of free time I couldn't able to further analyze the data , but according to this dataset the missing data can be further predicted by passing a set of variables to the machine learning algorithm like a regressor or a classifier (or) some missing data can be neglected based on the priority needed, Which I would like to continue if you can provide me some more time (probably a week).