Homework 2

Due Date: 5/30/2021

Data Definition:

The dataset used for our analysis was the Dunkin Donuts data. The URL for the dataset is posted below. It is comprised of details for individual Dunkin Donuts stores across the United States, such as their location, daily hours, services offered (drive thru, delivery, etc), and much more.

Data Exploration:

We first loaded the data from the URL and created a dictionary object to read the JSON data. From there we were able to convert the dictionary object into a pandas data frame. Once the data frame was created, we went through the following exploration and cleansing steps:

1. View the data frame to gain an understanding of what it contained
2. Deleted the unnecessary columns
3. Ensured all column headers were present
4. Created a copy data frame to perform analysis on

Question 1: Which state has the most dunkin donuts locations that offers mobile service?

The first step we took to answer this question was to create a subset data frame, pulling in only the columns necessary to answer the question, .mobile and .state. To investigate the new data frame, we created a pivot table, grouping by .mobile to identify which states contained stores that offered mobile services.

.mobile	.state
N	AK
	AL
	AR
	AZ
	CA
...	...
Y	VA
	VT
	WI
	WV
	WY

Viewing the data in this format made it difficult to identify the number of stores per state and how many offered mobile services, so we utilized the get.dummies function to turn the mobile variable into a

binary variable. Once the variables were created, we concatenated the two data frames and created a pivot table - grouping by mobile and state, calculated the sum of Y, then sorted in descending order.

		Y	sum
.mobile	.state		
Y	NY	1340.0	NaN
	MA	1053.0	NaN
	NJ	832.0	NaN
	FL	822.0	NaN
	IL	668.0	NaN
...	...	...	...
N	NJ	0.0	NaN
	NH	0.0	NaN
	NE	0.0	NaN
	NC	0.0	NaN
	AK	0.0	NaN

The conclusion that we were able to draw from this table is that NY has the most Dunkin Donuts locations at 1,340 that offers mobile service.

Question 2: Investigate the state and city with the highest number of sites offering delivery through GrubHub, Door Dash, and Uber Eats. What is the average number of delivery options for those sites?

To answer this question, we first created a data frame containing the state, city, and activity columns for each delivery method. Once that was completed, we created two calculated columns – the first reflecting the number of stores per city, and the second reflecting the number of delivery methods available at each site.

	.state	.city	.gh_active	.dd_active	.ue_active	City_Sites	Delivery_Methods
0	CA	South San Francisco	Y	Y	Y	1	3
1	CA	Oakland	Y	Y	Y	3	3
2	CA	Hayward	Y	Y	Y	1	3
3	CA	San Carlos	Y	Y	Y	1	3
4	CA	Half Moon Bay		Y	Y	1	2

This allowed us to create a consolidated data frame, which grouped the sites together by their city and state and calculated the number of sites in the city and total number of delivery methods. From there, we calculated the average number of delivery methods available per site in each city. We sorted the data frame by the total number of sites and created a data frame containing the top 10 cities.

.city	.state	City_Sites	Delivery_Methods	avg_delivery_methods
Chicago	IL	214	465	2.172897
New York	NY	166	423	2.548193
Brooklyn	NY	139	338	2.431655
Philadelphia	PA	131	338	2.580153
Bronx	NY	94	257	2.734043
Boston	MA	70	98	1.400000
Orlando	FL	49	106	2.163265
Jamaica	NY	41	76	1.853659
Baltimore	MD	40	101	2.525000
Staten Island	NY	39	109	2.794872

We then sorted the top 10 cities by the average number of delivery methods. This allowed us to clearly see which cities had the highest number of sites and how the number of sites related to the number of delivery methods available.

.city	.state	City_Sites	Delivery_Methods	avg_delivery_methods
Staten Island	NY	39	109	2.794872
Bronx	NY	94	257	2.734043
Philadelphia	PA	131	338	2.580153
New York	NY	166	423	2.548193
Baltimore	MD	40	101	2.525000
Brooklyn	NY	139	338	2.431655
Chicago	IL	214	465	2.172897
Orlando	FL	49	106	2.163265
Jamaica	NY	41	76	1.853659
Boston	MA	70	98	1.400000

These two visuals enabled us to draw the conclusion that Chicago, Illinois ranks the highest in the city with the most sites. Of their 214 sites, they have an average of 2.17 delivery methods per site. Although Chicago has the highest number of sites, they rank 7th amongst the top 10 in the number of delivery methods per site. Staten Island has the highest volume of sites which offer all three delivery methods.