Programming Assignment 2

Name: Jimmy Nguyen Date: October 9th, 2022 Course/Professor: INST447/Sigalo

- Research Question: Do people order different foods at different times of day?

In [10]:
```python
import matplotlib.pyplot as plt
%matplotlib inline
import pandas as pd
```

In [13]:
```python
df = pd.read_csv('InstacartOrdersByDepartment.csv', encoding = 'unicode_escape')
df
```

Out[13]:

	order_hour_of_day	department	num_orders_hour	tot_orders_dept
0	0	alcohol	33.0	5598.0
1	1	alcohol	32.0	5598.0
2	2	alcohol	5.0	5598.0
3	3	alcohol	3.0	5598.0
4	4	alcohol	2.0	5598.0
...	...	...	...	...
499	19	snacks	4959.0	118862.0
500	20	snacks	3397.0	118862.0
501	21	snacks	2782.0	118862.0
502	22	snacks	2216.0	118862.0
503	23	snacks	1322.0	118862.0

504 rows × 4 columns

Visualization 1

In [45]:
```python
df_alc = df[0:24]
df_alc.head()
```

Out[45]:

	order_hour_of_day	department	num_orders_hour	tot_orders_dept
0	0	alcohol	33.0	5598.0
1	1	alcohol	32.0	5598.0
2	2	alcohol	5.0	5598.0
3	3	alcohol	3.0	5598.0
4	4	alcohol	2.0	5598.0

```
In [68]:   fig = plt.figure(figsize = (10, 4))
           plt.bar(df_alc['order_hour_of_day'], df_alc['num_orders_hour'])
           plt.xlabel("Orders Per Hour Of Day")
           plt.ylabel("Number of Order Per Hours")
           plt.title("Number of Orders of Alcohol Throughout A Day")
           plt.show()
```


Visualization 1: Description of the Results

From the visual data above, we can see that the distribution curve starts around 9AM till around 6PM. This would make sense because during the day, the general public is up and about. Another observation that can be made is that the graph seems to spike around 3PM. This pattern makes sense as "Happy Hour" occurs around 3PM and many co-workers are looking to drink and hang out after work.
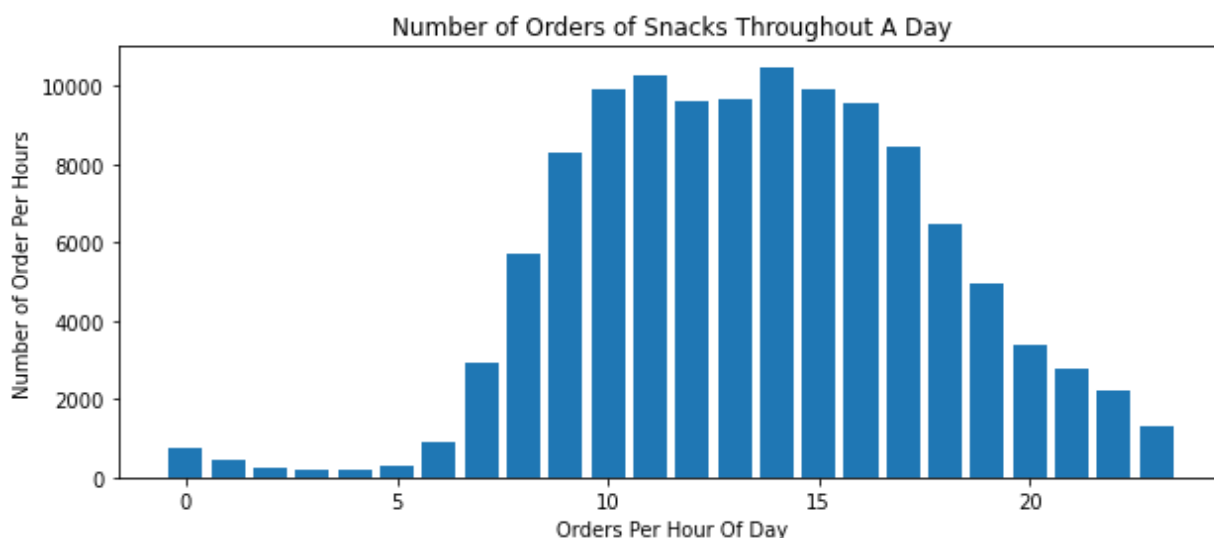
Visualization 2

```
In [54]:   df_sna = df[480:504]
           df_sna.head()
```

Out[54]:

	order_hour_of_day	department	num_orders_hour	tot_orders_dept
480	0	snacks	759.0	118862.0
481	1	snacks	455.0	118862.0
482	2	snacks	244.0	118862.0
483	3	snacks	203.0	118862.0
484	4	snacks	194.0	118862.0

```
In [67]:   fig = plt.figure(figsize = (10, 4))
           plt.bar(df_sna['order_hour_of_day'], df_sna['num_orders_hour'])
           plt.xlabel("Orders Per Hour Of Day")
           plt.ylabel("Number of Order Per Hours")
```

```
plt.title("Number of Orders of Snacks Throughout A Day")
plt.show()
```


Visualization 2: Description of the Results

From this graph above, we can tell that snacks purchases peak around 10AM-11AM and 2-3PM.
These purchases are most likely made because lunch and early dinners occur during this time. In
addition, I noticed that night purchases occur more drastically than the previous graph above.
We can deduct that movie nights are happening at this time (along with other night events) and
lots of people gravitate towards buying snacks versus alcoholic beverages at night.

Visualization 3

In [65]:
```
df_brk = df[96:120]
df_brk.head()
```

Out[65]:

	order_hour_of_day	department	num_orders_hour	tot_orders_dept
96	0	breakfast	175.0	29500.0
97	1	breakfast	124.0	29500.0
98	2	breakfast	60.0	29500.0
99	3	breakfast	54.0	29500.0
100	4	breakfast	62.0	29500.0

In [69]:
```
fig = plt.figure(figsize = (10, 4))
plt.bar(df_brk['order_hour_of_day'], df_brk['num_orders_hour'])
plt.xlabel("Orders Per Hour Of Day")
plt.ylabel("Number of Order Per Hours")
plt.title("Number of Orders of Breakfast Throughout A Day")
plt.show()
```

Number of Orders of Breakfast Throughout A Day
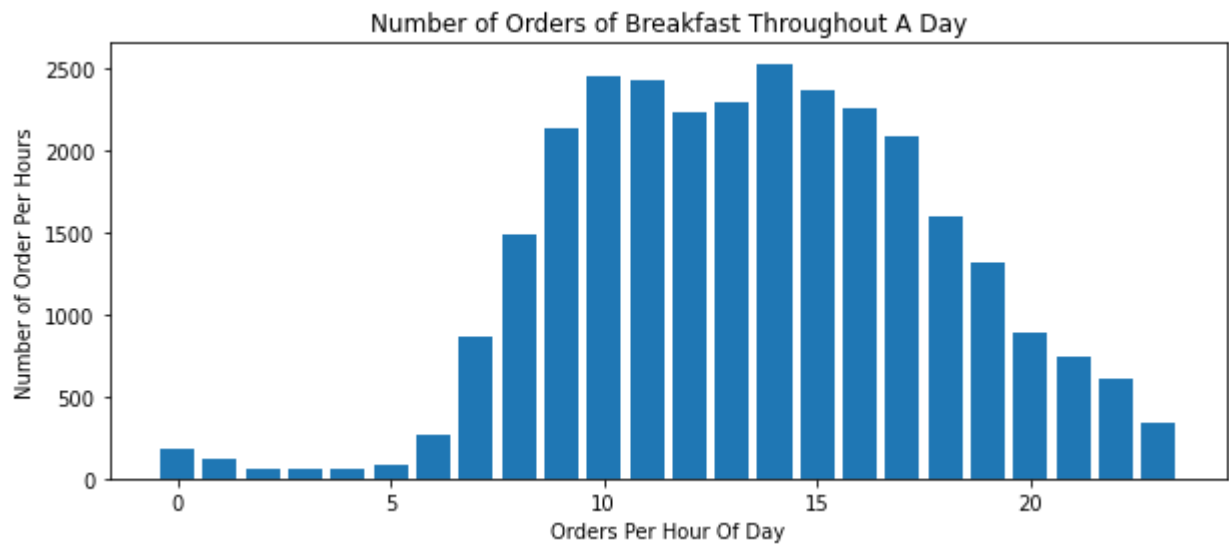
Visualiation 3: Description of the Results

From the database above, we see that breakfast quickly picks up in the early morning and peaks around 10AM-11AM. It makes sense though as many people are having their breakfast during this time. Additionally, I noticed that it also peaks around 2PM as well, which is very interesting. Upon further speculation, I realized that there are late bloomers who are having brunch with her friends or families throughout this time of the day so it makes sense.

Conclusion

To summarize, the similarities between all three graphs is that the orders begin rising around 5AM and end around 12PM. 'Alcoholic' orders peaks around 3PM because many co-workers are looking forward to "Happy Hour" but the purchases quickly die out before 12 midnight. 'Snacks' and 'Breakfast' orders both peak around 10AM and 2PM as those times directly correlate to breakfast and lunch. The main difference between these two graphs are that 'Snack' orders are still being made late into the night as I have hypothesized that people are looking for a quick snack for movie nights and other late night events. 'Breakfast' peaks at 2PM are presumed to be because there are those who wake up late or have decided to a late brunch with their friends/family.

The limitations of this dataset for the analyses is that we don't have information on which day of the week that these situations are occurring. I believe that if we further investigate the orders prior to the day of the week, there would be more informative assumptions that could be made based on that information alone. For example, we could find that 'alcoholic' orders occur on Friday nights most often then on the rest of the week.

In [ ]: