In [ ]:
```
import numpy as np
import matplotlib as mpl
import matplotlib.pyplot as plt
import pandas as pd
```

In [ ]:
```
url = 'https://raw.githubusercontent.com/justmarkham/DAT8/master/data/chipotle.tsv'
```

In [ ]:
```
df = pd.read_csv(url, sep = '\t',low_memory = False)
df.head()
```

In [ ]:
```
# print total records and type of variables
```

In [ ]:
```
df.info()
```

In [ ]:
```
#Q1: Which was the most ordered item? and How many items were ordered?
```

In [ ]:
```
c = df.groupby("item_name").sum()
c = c.sort_values(["quantity"], ascending = False)
c.head()
```

In [ ]:
```
#Q2: What was the most ordered item in the choice_description column?
```

In [ ]:
```
df = df.groupby("choice_description").sum()
df = df.sort_values(["quantity"], ascending = False)
df.head(1)
```

In [ ]:
```
#Q3: Turn the item price into a float
```

In [ ]:
```
dollar = lambda x: float(x[1:-1])
df.item_price = df.item_price.apply(dollar)
```

In [ ]:
```
df.head()
```

In [ ]:
```
#Q3: How much was the revenue for the period in the dataset?
```

In [ ]:
```
revenue = (df['quantity']* df['item_price']).sum()
print('Revenue was: ' + str(np.round(revenue,2)))
```

In [ ]:
```
#Q4: print a data frame with only two columns item_name and item_price
```

In [ ]:
```
df.loc[:,["item_name","item_price"]]
```

In [ ]:
```
#Q5: delete the duplicates in item_name and quantity
```

In [ ]:
```
filtered = df.drop_duplicates(['item_name','quantity'])
# select only the products with quantity equals to 1
one_prod = filtered[filtered.quantity == 1]
# select only the item_name and item_price columns
price_per_item = one_prod[['item_name', 'item_price']]
# sort the values from the most to more expensive
price_per_item.sort_values(by = "item_price", ascending = True)
```

In [ ]:
```
#Q6: What was the quantity of the most expensive item ordered?
```

In [ ]:
```
df.sort_values(["item_price"],ascending = False).head(1)
```

In [ ]:
```
# Q7: How many times were a Veggie Salad Bowl ordered?
```

In [ ]:
```
df = df[df.item_name == "Veggie Salad Bowl"]
count_row = df.shape[0]
print(count_row)
```

Drinks statistics

In [47]:
```
url = "https://raw.githubusercontent.com/alcor2019/justmarkham/master/data/drinks.csv"
```

In [49]:
```
df = pd.read_csv(url,low_memory = False)
df.head()
```

Out[49]:

	country	beer_servings	spirit_servings	wine_servings	total_litres_of_pure_alcohol	continent
0	Afghanistan	0	0	0	0.0	Asia
1	Albania	89	132	54	4.9	Europe
2	Algeria	25	0	14	0.7	Africa
3	Andorra	245	138	312	12.4	Europe
4	Angola	217	57	45	5.9	Africa

In [50]:
```
#Q8: Which continent drinks more beer on average?
```

In [51]:
```
x = df.groupby("continent")
x.beer_servings.mean()
```

Out[51]:
```
continent
Africa            61.471698
Asia              37.045455
Europe           193.777778
North America    145.434783
Oceania           89.687500
South America    175.083333
Name: beer_servings, dtype: float64
```

In [52]:
```
#Q9: For each continent print the statistics for wine consumption.
```

In [53]:
```
df.groupby("continent").wine_servings.describe()
```

Out[53]:

continent	count	mean	std	min	25%	50%	75%	max
Africa	53.0	16.264151	38.846419	0.0	1.0	2.0	13.00	233.0
Asia	44.0	9.068182	21.667034	0.0	0.0	1.0	8.00	123.0
Europe	45.0	142.222222	97.421738	0.0	59.0	128.0	195.00	370.0
North America	23.0	24.521739	28.266378	1.0	5.0	11.0	34.00	100.0
Oceania	16.0	35.625000	64.555790	0.0	1.0	8.5	23.25	212.0
South America	12.0	62.416667	88.620189	1.0	3.0	12.0	98.50	221.0

In [ ]:
```

```