Chipotle Pandas Project

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

In [ ]:
```
data=pd.read_csv('chipotle.tsv',sep='\t')
```

DATA CLEANING

viewing first 10 rows

In [ ]:
```
data.head(10)
```

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

remvoving dollar sign and converitng to float

In [ ]:
```
# removing dollar sign
data['item_price']=data['item_price'].str[1:]
```

In [ ]:
```
# converting to float
data['item_price']=data['item_price'].astype('float')
```

checking for null values

In [ ]:
```
data.isnull().sum()
```

shape of dataset

In [ ]:
```
data.shape
```

formatting the choice_description column

In [ ]:
```
# replacing null value
data.fillna("no description", inplace=True)
```

In [ ]:
```
data['choice_description']=data['choice_description'].apply(lambda x: x.rep
```

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

Analysis

most orderd item

In [ ]:
```
#creating dataframe with items and quantity orderd
item_quantity=data.groupby('item_name')['quantity'].sum()
item_quantity_sort=pd.DataFrame(item_quantity.sort_values(ascending=False))
item_quantity_sort.head()
```

In [ ]:
```
# most orderd item
item_quantity_sort.iloc[0]
```

top 10 best selling items

In [ ]:
```
top_10=pd.DataFrame(item_quantity_sort.iloc[0:10])
top_10.plot(kind='bar')
```

most ordered item in choice_description

In [ ]:
```
choice_quantity = data.groupby('choice_description')['quantity'].sum()
choice_quantity_sorted = pd.DataFrame(choice_quantity.sort_values(ascending
choice_quantity_sorted.iloc[0]
```

Total orders

In [ ]:
```
orders=data.groupby('order_id')
len(orders)
```

Average number of items in an order

In [ ]:
```
order_quantity=data.groupby('order_id')['quantity'].sum()
order_quantity_df=pd.DataFrame(order_quantity)
order_quantity_df['quantity'].mean()
```

Number of items sold

In [ ]:
```
data['item_name'].nunique()
```

Total revenue

In [ ]:
```
# adding revenue to data
data["revenue"]=data['quantity']*data['item_price']
```

In [ ]:
```
data['revenue'].sum()
```

Average revenue per order

In [ ]:
```
order_group=data.groupby('order_id')['revenue'].sum()
order_group.mean()
```

Number of Soft drinks Orderd

In [ ]:
```
soft_drink=data.groupby('item_name')
soft_drink_df=soft_drink.get_group('Canned Soft Drink')
soft_drink_df['quantity'].sum()
```

Soft drink preferance

In [ ]:
```
soft_drink_quant=soft_drink_df[['quantity','choice_description']]
sf_orderd=soft_drink_quant.groupby('choice_description')['quantity'].sum()
sf_orderd
```

In [ ]:
```
# visulasing using pie chart
sf_orderd.plot(kind='pie')
```

Salsa Type Preference of customers

In [ ]:
```
# creating a dataframe with orders of salsa
salsa = data.copy()
```

In [ ]:
```
# editing data frame to have only type of salsa ordered
salsa['choice_description']=data['choice_description'].apply(lambda x : x.s
salsa=salsa[salsa['choice_description'].str.contains('Salsa')]
```

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

In [ ]:
```
# finding quantity of each salsa
salsa_quantity=salsa.groupby('choice_description')['quantity'].sum()
salsa_quantity
```

In [ ]:
```
salsa_quantity.plot(kind="pie")
```

In [ ]:
```

```