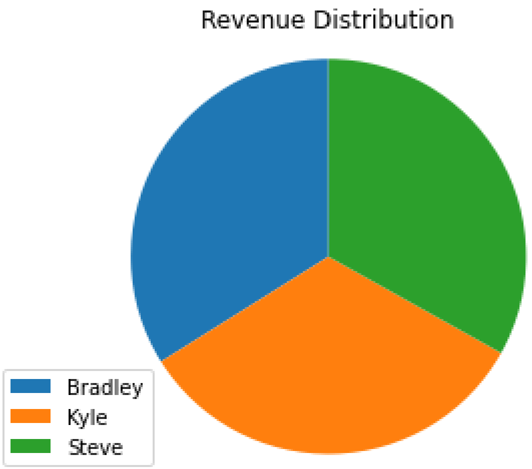
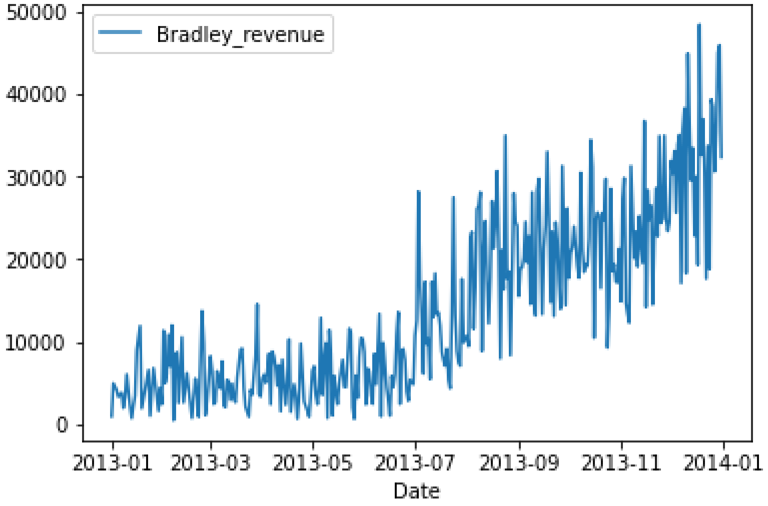
In [9]:

```
import pandas as pd
import numpy as np
import random
import matplotlib.pyplot as plt
import pandasql as ps

#Read sample sales data from CSV

df = pd.read_csv('https://raw.githubusercontent.com/ine-rmotr-curriculum/FreeCodeCamp-P
                 header = [0],
                 parse_dates=['Date'],
                 skip_blank_lines=True)


#Only select rows with a not null Revenue value


df = df[df['Revenue'].notna()]


#Create list for salesperson id and add a random value for every record in the sales da
#Then transpose list to a column and add it to the dataframe


Salesperson_id = []

for x in np.arange(df.shape[0]):
    Salesperson_id.append(random.randint(1,3))

df['Salesperson_id'] = Salesperson_id


#Create dictionary for each salesperson_id and their corresponding name,
#Convert the dictionary to a dataframe, then join it to the original dataframe using th


d = {'Salesperson_id': [1, 2, 3],
    'Salesperson_nm' : ['Bradley','Steve','Kyle']}
df2 = pd.DataFrame(data=d)
df3 = pd.merge(df,df2,on='Salesperson_id')


#Panda SQL query to find revenue per salesperson, then plot the data in a pie chart



df4 = ps.sqldf("select Salesperson_nm,SUM(Revenue) as Revenue from df3 group by Salespe
df4.set_index('Salesperson_nm')


labels = df4['Salesperson_nm']
revs = df4['Revenue']


patches, texts = plt.pie(revs, startangle=90)
plt.legend(patches, labels, loc="best")
plt.title("Revenue Distribution")
plt.axis('equal')
```

```
#Panda SQL queries to find out revenue of each salesperson per day in the year 2013
#Then merge the query results into 1 dataframe


df5 = ps.sqldf("select Date,SUM(Revenue) as Bradley_revenue from df3 where Salesperson_
df6 = ps.sqldf("select Date,SUM(Revenue) as Steve_revenue from df3 where Salesperson_nm
df7 = ps.sqldf("select Date,SUM(Revenue) as Kyle_revenue from df3 where Salesperson_nm


merged1 = pd.merge(df5,df6,on="Date")
merged2 = pd.merge(merged1,df7,on="Date")

merged2['Date'] = pd.to_datetime(merged2['Date']).dt.date
merged2.set_index('Date')

#merged2.head()

#Divide outlier daily revenue values for Bradley by 3

merged2.loc[merged2['Bradley_revenue'] > 50000, 'Bradley_revenue'] = merged2.loc[merged


#Plot Bradley's revenue in 2013

merged2.plot(x='Date', y='Bradley_revenue');
```





In [ ]: