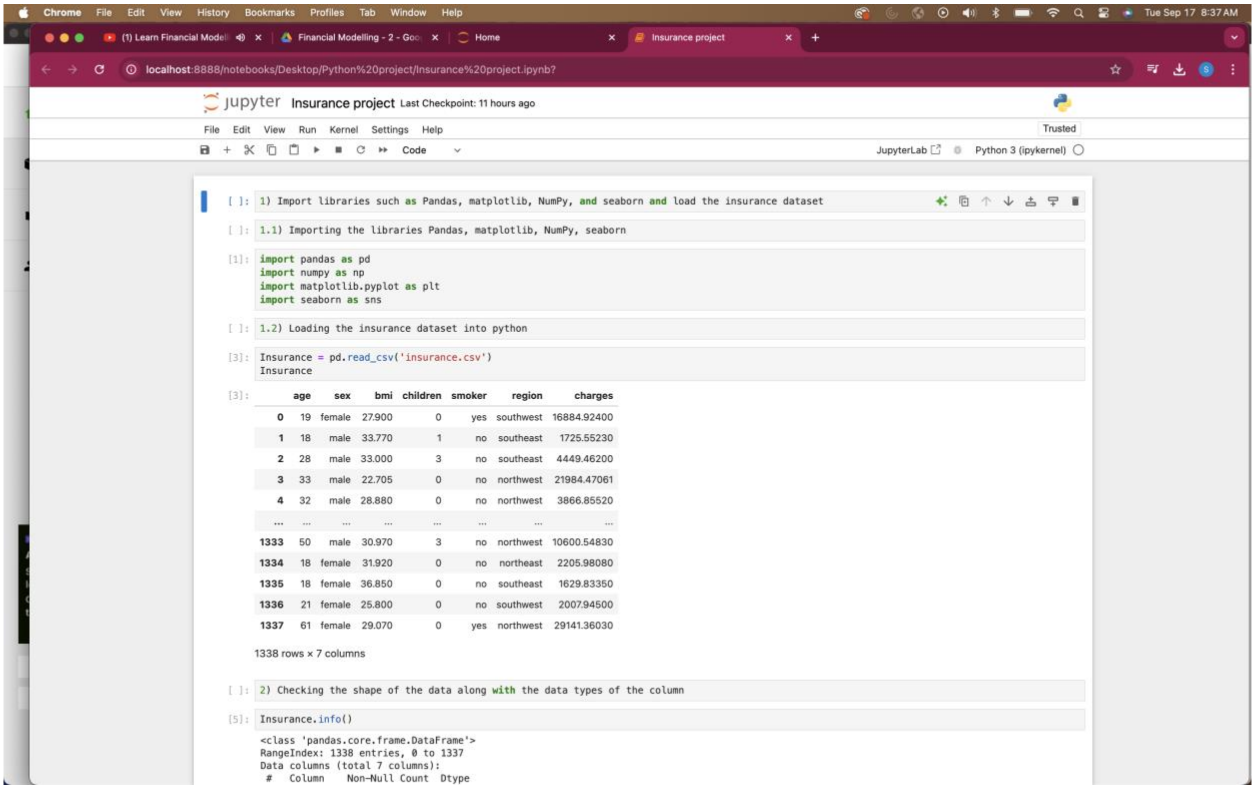
Insurance project


Chrome File Edit View History Bookmarks Profiles Tab Window Help
Tue Sep 17 8:37 AM
(1) Learn Financial Modell
Financial Modelling - 2 - Goo
Home
Insurance project
localhost:8888/notebooks/Desktop/Python%20project/Insurance%20project.ipynb?
jupyter Insurance project Last Checkpoint: 11 hours ago
File Edit View Run Kernel Settings Help
Trusted
Code
JupyterLab
Python 3 (ipykernel)

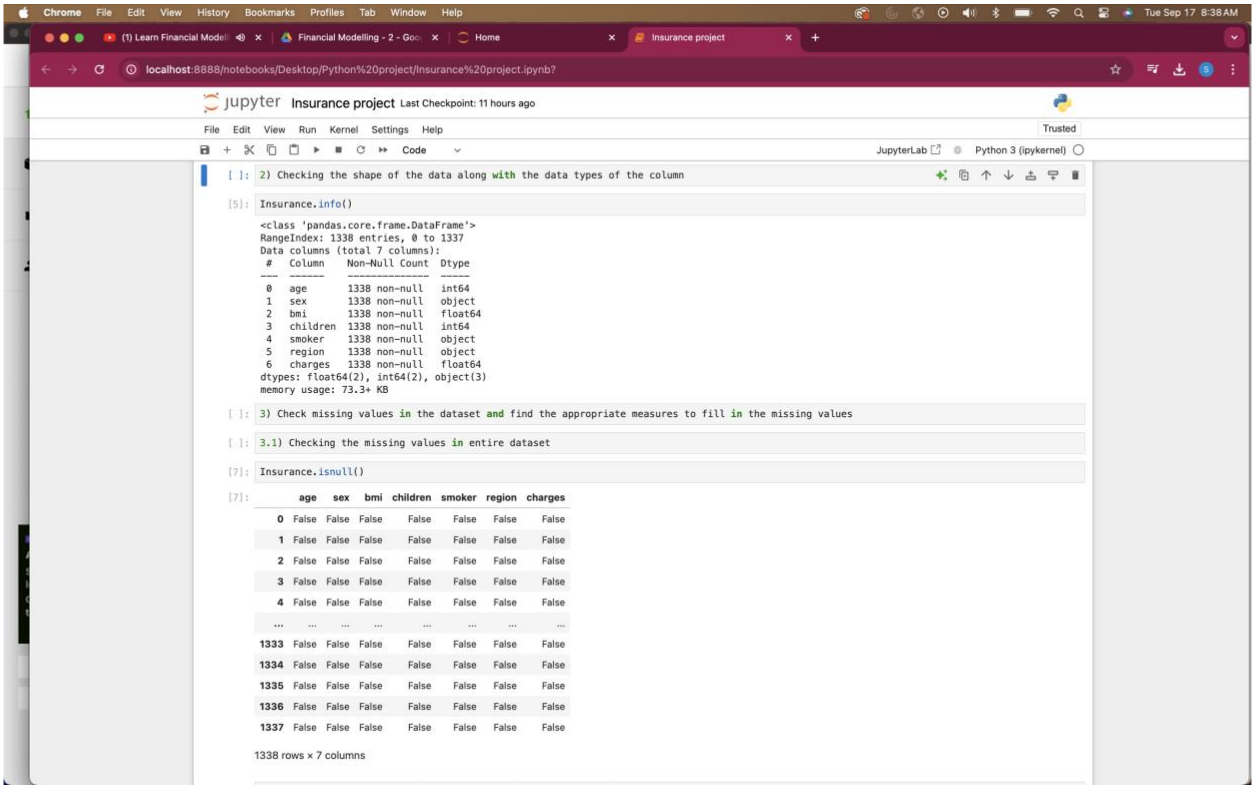
```
[ ]: 1) Import libraries such as Pandas, matplotlib, NumPy, and seaborn and load the insurance dataset

[ ]: 1.1) Importing the libraries Pandas, matplotlib, NumPy, seaborn

[1]: import pandas as pd
     import numpy as np
     import matplotlib.pyplot as plt
     import seaborn as sns

[ ]: 1.2) Loading the insurance dataset into python

[3]: Insurance = pd.read_csv('insurance.csv')
     Insurance
```

[3]:

	age	sex	bmi	children	smoker	region	charges
0	19	female	27.900	0	yes	southwest	16884.92400
1	18	male	33.770	1	no	southeast	1725.55230
2	28	male	33.000	3	no	southeast	4449.46200
3	33	male	22.705	0	no	northwest	21984.47061
4	32	male	28.880	0	no	northwest	3866.85520
...	...	...	...	...	...	...	...
1333	50	male	30.970	3	no	northwest	10600.54830
1334	18	female	31.920	0	no	northeast	2205.98080
1335	18	female	36.850	0	no	southeast	1629.83350
1336	21	female	25.800	0	no	southwest	2007.94500
1337	61	female	29.070	0	yes	northwest	29141.36030

1338 rows × 7 columns

```
[ ]: 2) Checking the shape of the data along with the data types of the column

[5]: Insurance.info()

     <class 'pandas.core.frame.DataFrame'>
     RangeIndex: 1338 entries, 0 to 1337
     Data columns (total 7 columns):
      #   Column    Non-Null Count  Dtype
```

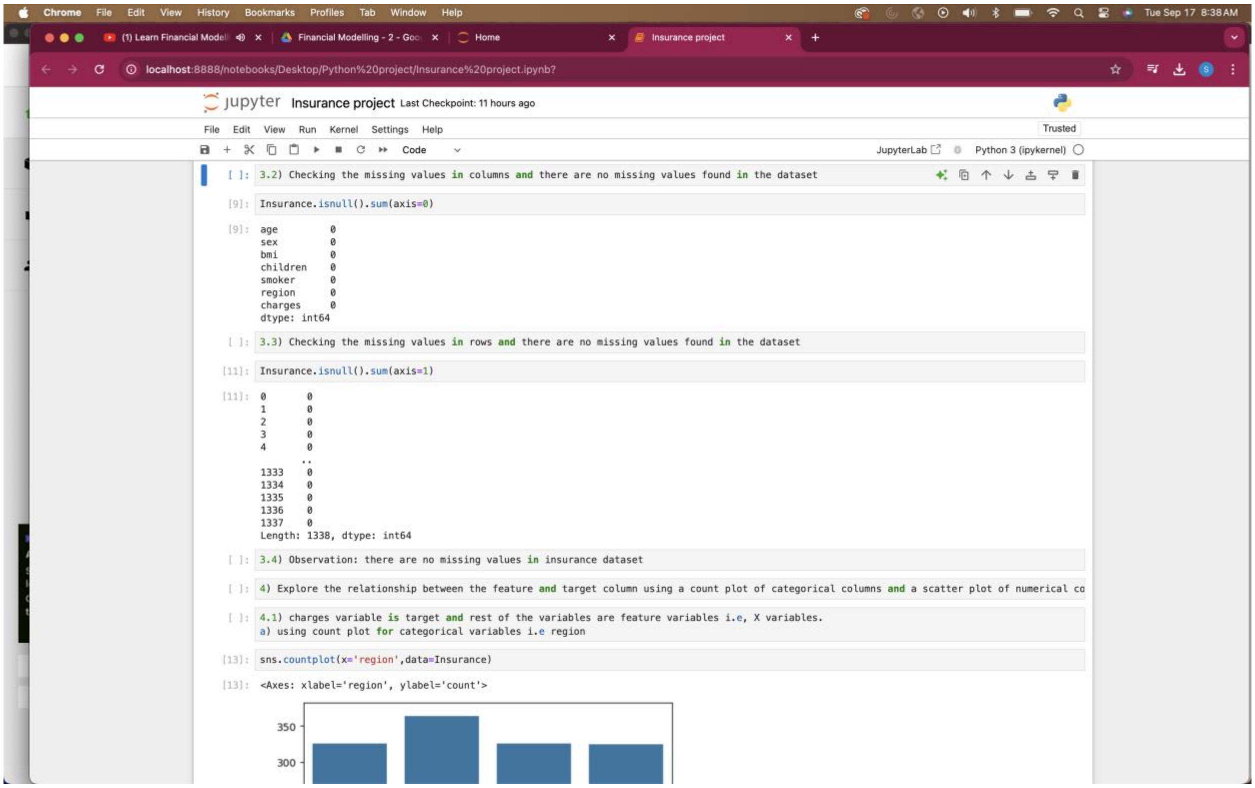

Chrome File Edit View History Bookmarks Profiles Tab Window Help
Tue Sep 17 8:38 AM
(1) Learn Financial Modell
Financial Modelling - 2 - Goo
Home
Insurance project
localhost:8888/notebooks/Desktop/Python%20project/Insurance%20project.ipynb?
jupyter Insurance project Last Checkpoint: 11 hours ago
File Edit View Run Kernel Settings Help
Trusted
Code
JupyterLab
Python 3 (ipykernel)

```
[ ]: 2) Checking the shape of the data along with the data types of the column

[5]: Insurance.info()

     <class 'pandas.core.frame.DataFrame'>
     RangeIndex: 1338 entries, 0 to 1337
     Data columns (total 7 columns):
      #   Column    Non-Null Count  Dtype
     ---  ------    --------------  -----
      0   age       1338 non-null   int64
      1   sex       1338 non-null   object
      2   bmi       1338 non-null   float64
      3   children  1338 non-null   int64
      4   smoker    1338 non-null   object
      5   region    1338 non-null   object
      6   charges   1338 non-null   float64
     dtypes: float64(2), int64(2), object(3)
     memory usage: 73.3+ KB

[ ]: 3) Check missing values in the dataset and find the appropriate measures to fill in the missing values

[ ]: 3.1) Checking the missing values in entire dataset

[7]: Insurance.isnull()
```

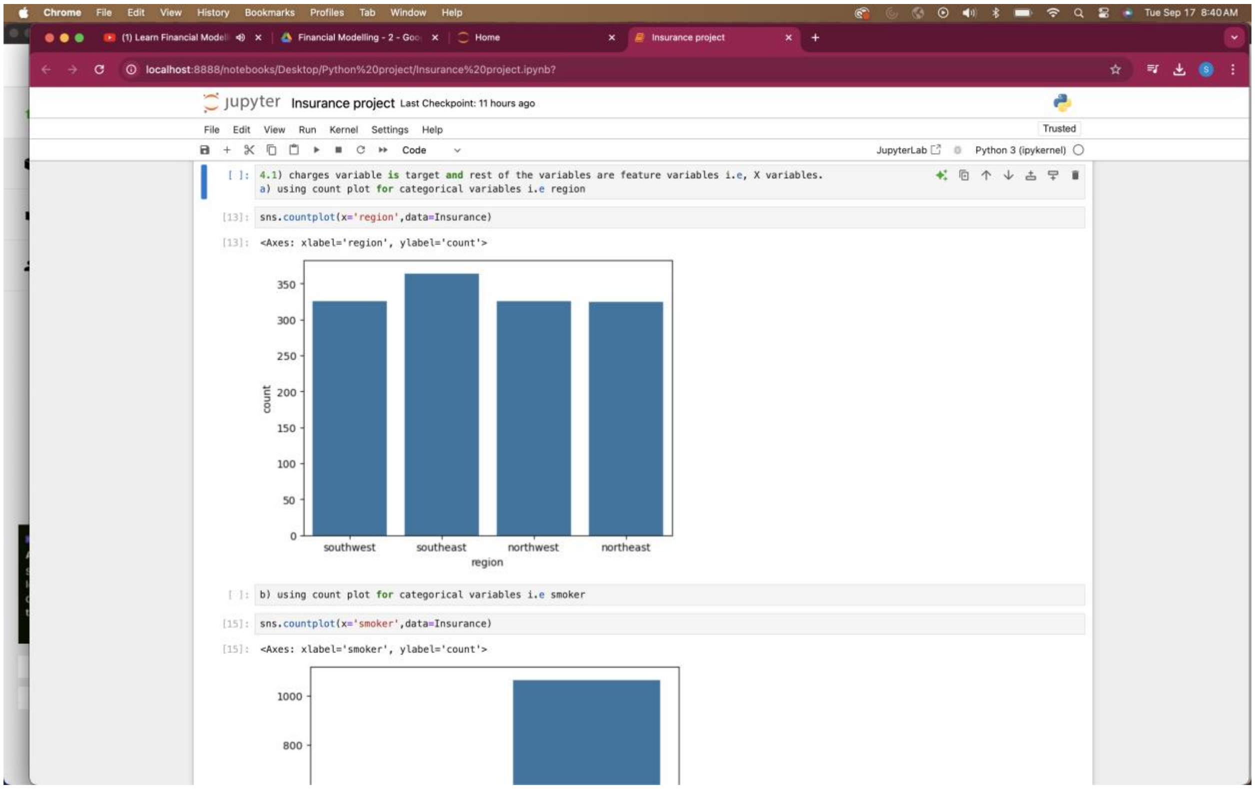
[7]:

	age	sex	bmi	children	smoker	region	charges
0	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...
1333	False	False	False	False	False	False	False
1334	False	False	False	False	False	False	False
1335	False	False	False	False	False	False	False
1336	False	False	False	False	False	False	False
1337	False	False	False	False	False	False	False

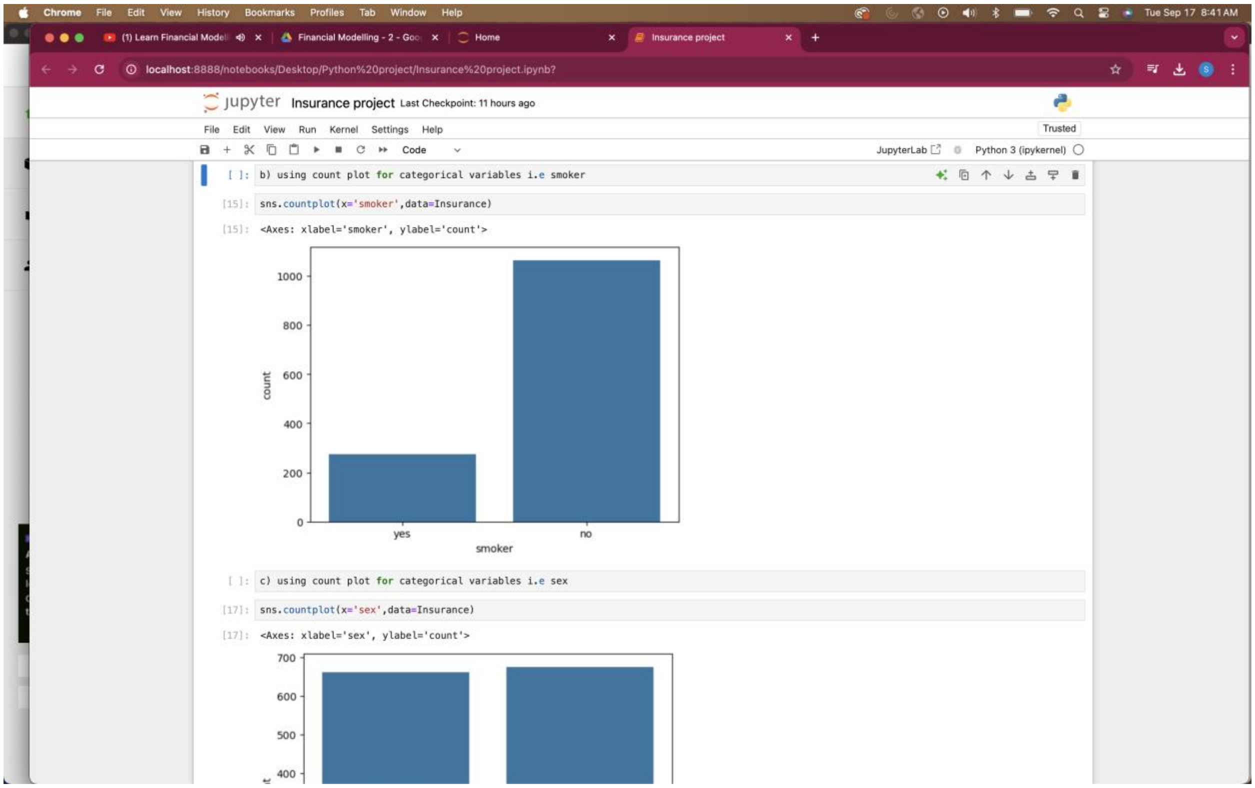
1338 rows × 7 columns


Chrome File Edit View History Bookmarks Profiles Tab Window Help
Tue Sep 17 8:38 AM
(1) Learn Financial Modeli
Financial Modelling - 2 - Goo
Home
Insurance project
localhost:8888/notebooks/Desktop/Python%20project/Insurance%20project.ipynb?
jupyter
Insurance project
Last Checkpoint: 11 hours ago
File Edit View Run Kernel Settings Help
Trusted
Code
JupyterLab
Python 3 (ipykernel)
350
300

```
[ ]: 3.2) Checking the missing values in columns and there are no missing values found in the dataset

[9]: Insurance.isnull().sum(axis=0)

[9]: age         0
     sex         0
     bmi         0
     children    0
     smoker      0
     region      0
     charges     0
     dtype: int64

[ ]: 3.3) Checking the missing values in rows and there are no missing values found in the dataset

[11]: Insurance.isnull().sum(axis=1)

[11]: 0       0
      1       0
      2       0
      3       0
      4       0
             ..
      1333    0
      1334    0
      1335    0
      1336    0
      1337    0
      Length: 1338, dtype: int64

[ ]: 3.4) Observation: there are no missing values in insurance dataset

[ ]: 4) Explore the relationship between the feature and target column using a count plot of categorical columns and a scatter plot of numerical co

[ ]: 4.1) charges variable is target and rest of the variables are feature variables i.e, X variables.
     a) using count plot for categorical variables i.e region

[13]: sns.countplot(x='region',data=Insurance)

[13]: <Axes: xlabel='region', ylabel='count'>
```

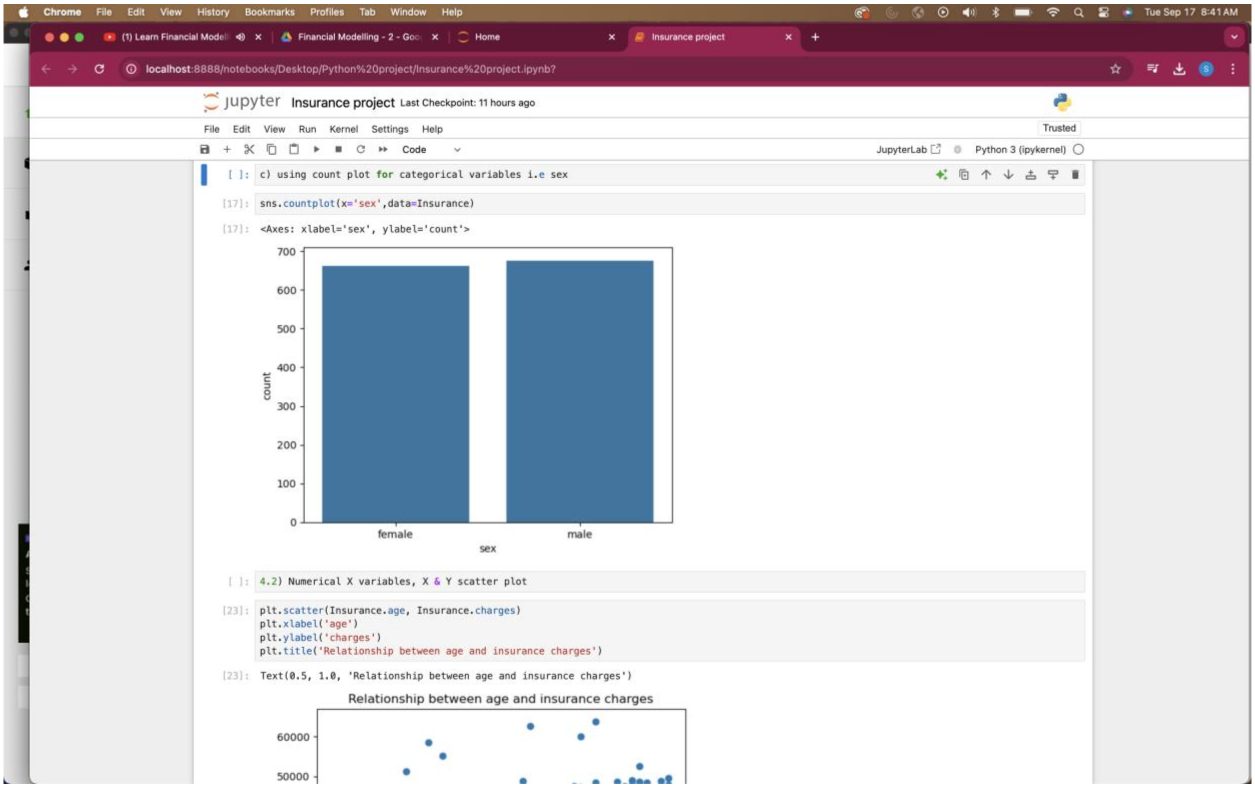

Chrome File Edit View History Bookmarks Profiles Tab Window Help
Tue Sep 17 8:40 AM
(1) Learn Financial Modeli
Financial Modelling - 2 - Goo
Home
Insurance project
localhost:8888/notebooks/Desktop/Python%20project/Insurance%20project.ipynb?
jupyter
Insurance project
Last Checkpoint: 11 hours ago
File Edit View Run Kernel Settings Help
Trusted
Code
JupyterLab
Python 3 (ipykernel)
350
300
250
200
150
100
50
0
count
southwest
southeast
northwest
northeast
region
1000
800

```
[ ]: 4.1) charges variable is target and rest of the variables are feature variables i.e, X variables.
     a) using count plot for categorical variables i.e region

[13]: sns.countplot(x='region',data=Insurance)

[13]: <Axes: xlabel='region', ylabel='count'>

[ ]: b) using count plot for categorical variables i.e smoker

[15]: sns.countplot(x='smoker',data=Insurance)

[15]: <Axes: xlabel='smoker', ylabel='count'>
```


Chrome File Edit View History Bookmarks Profiles Tab Window Help
Tue Sep 17 8:41 AM
(1) Learn Financial Modell
Financial Modelling - 2 - Goo
Home
Insurance project
localhost:8888/notebooks/Desktop/Python%20project/Insurance%20project.ipynb?
jupyter Insurance project Last Checkpoint: 11 hours ago
File Edit View Run Kernel Settings Help
Trusted
Code
JupyterLab
Python 3 (ipykernel)
1000
800
600
400
200
0
count
yes
no
smoker
700
600
500
400

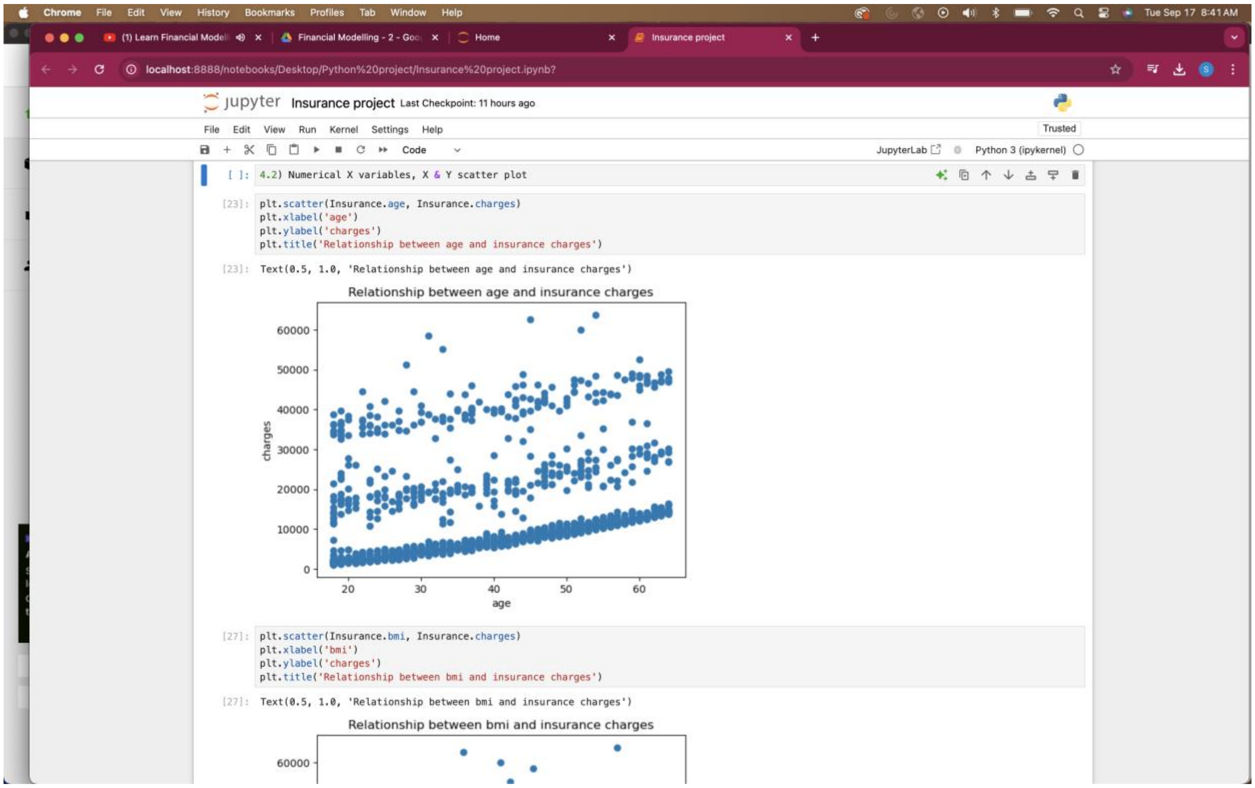
```
[ ]: b) using count plot for categorical variables i.e smoker

[15]: sns.countplot(x='smoker',data=Insurance)

[15]: <Axes: xlabel='smoker', ylabel='count'>
```

```
[ ]: c) using count plot for categorical variables i.e sex

[17]: sns.countplot(x='sex',data=Insurance)

[17]: <Axes: xlabel='sex', ylabel='count'>
```


Chrome File Edit View History Bookmarks Profiles Tab Window Help
Tue Sep 17 8:41 AM
(1) Learn Financial Modell
Financial Modelling - 2 - Goo
Home
Insurance project
localhost:8888/notebooks/Desktop/Python%20project/Insurance%20project.ipynb?
jupyter Insurance project Last Checkpoint: 11 hours ago
File Edit View Run Kernel Settings Help
Trusted
Code
JupyterLab
Python 3 (ipykernel)
700
600
500
400
300
200
100
0
count
female
male
sex
Relationship between age and insurance charges
60000
50000

```
[ ]: c) using count plot for categorical variables i.e sex

[17]: sns.countplot(x='sex',data=Insurance)

[17]: <Axes: xlabel='sex', ylabel='count'>
```

```
[ ]: 4.2) Numerical X variables, X & Y scatter plot

[23]: plt.scatter(Insurance.age, Insurance.charges)
      plt.xlabel('age')
      plt.ylabel('charges')
      plt.title('Relationship between age and insurance charges')

[23]: Text(0.5, 1.0, 'Relationship between age and insurance charges')
```


Chrome File Edit View History Bookmarks Profiles Tab Window Help
Tue Sep 17 8:41 AM
(1) Learn Financial Modell
Financial Modelling - 2 - Goo
Home
Insurance project
localhost:8888/notebooks/Desktop/Python%20project/Insurance%20project.ipynb?
jupyter Insurance project Last Checkpoint: 11 hours ago
File Edit View Run Kernel Settings Help
Trusted
Code
JupyterLab
Python 3 (ipykernel)

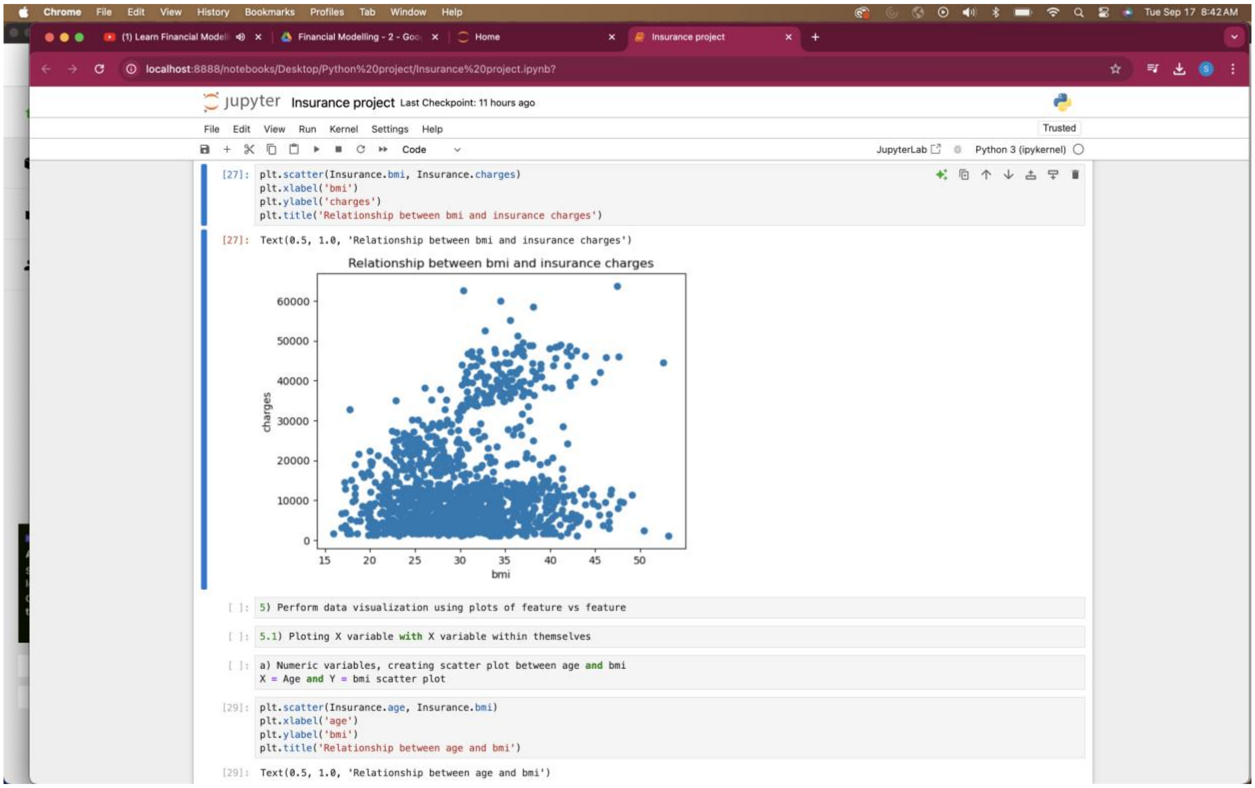
```
[ ]: 4.2) Numerical X variables, X & Y scatter plot

[23]: plt.scatter(Insurance.age, Insurance.charges)
      plt.xlabel('age')
      plt.ylabel('charges')
      plt.title('Relationship between age and insurance charges')

[23]: Text(0.5, 1.0, 'Relationship between age and insurance charges')
```

Relationship between age and insurance charges

60000
50000
40000
30000
20000
10000
0
charges
20
30
40
50
60
age

```
[27]: plt.scatter(Insurance.bmi, Insurance.charges)
      plt.xlabel('bmi')
      plt.ylabel('charges')
      plt.title('Relationship between bmi and insurance charges')

[27]: Text(0.5, 1.0, 'Relationship between bmi and insurance charges')
```

Relationship between bmi and insurance charges

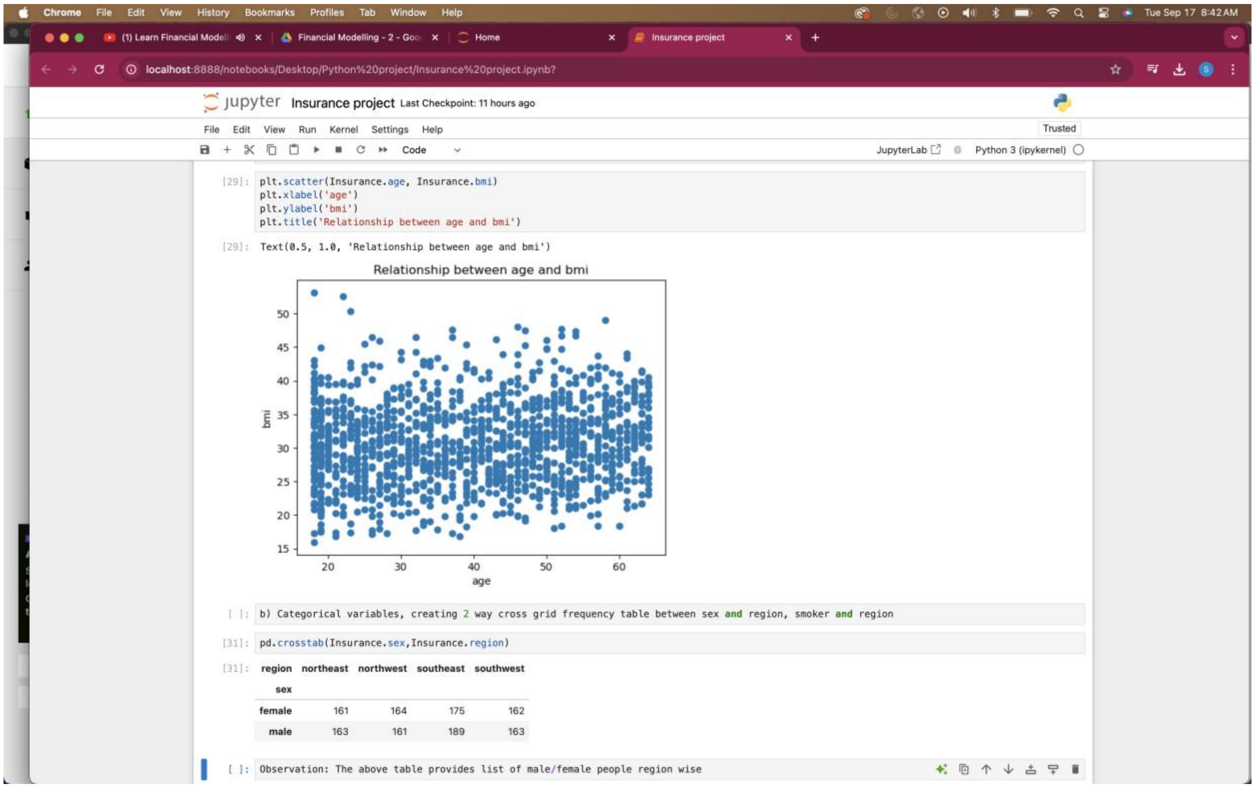

Chrome File Edit View History Bookmarks Profiles Tab Window Help
Tue Sep 17 8:42 AM
(1) Learn Financial Modell
Financial Modelling - 2 - Goo
Home
Insurance project
localhost:8888/notebooks/Desktop/Python%20project/Insurance%20project.ipynb?
jupyter Insurance project Last Checkpoint: 11 hours ago
File Edit View Run Kernel Settings Help
Trusted
Code
JupyterLab
Python 3 (ipykernel)

```
[27]: plt.scatter(Insurance.bmi, Insurance.charges)
      plt.xlabel('bmi')
      plt.ylabel('charges')
      plt.title('Relationship between bmi and insurance charges')

[27]: Text(0.5, 1.0, 'Relationship between bmi and insurance charges')
```

Relationship between bmi and insurance charges

60000
50000
40000
30000
20000
10000
0
charges
15
20
25
30
35
40
45
50
bmi

```
[ ]: 5) Perform data visualization using plots of feature vs feature

[ ]: 5.1) Ploting X variable with X variable within themselves

[ ]: a) Numeric variables, creating scatter plot between age and bmi
     X = Age and Y = bmi scatter plot

[29]: plt.scatter(Insurance.age, Insurance.bmi)
      plt.xlabel('age')
      plt.ylabel('bmi')
      plt.title('Relationship between age and bmi')

[29]: Text(0.5, 1.0, 'Relationship between age and bmi')
```


Chrome File Edit View History Bookmarks Profiles Tab Window Help
Tue Sep 17 8:42 AM
(1) Learn Financial Modell
Financial Modelling - 2 - Goo
Home
Insurance project
localhost:8888/notebooks/Desktop/Python%20project/Insurance%20project.ipynb?
jupyter Insurance project Last Checkpoint: 11 hours ago
File Edit View Run Kernel Settings Help
Trusted
Code
JupyterLab
Python 3 (ipykernel)
Relationship between age and bmi
50
45
40
35
30
25
20
15
bmi
20
30
40
50
60
age

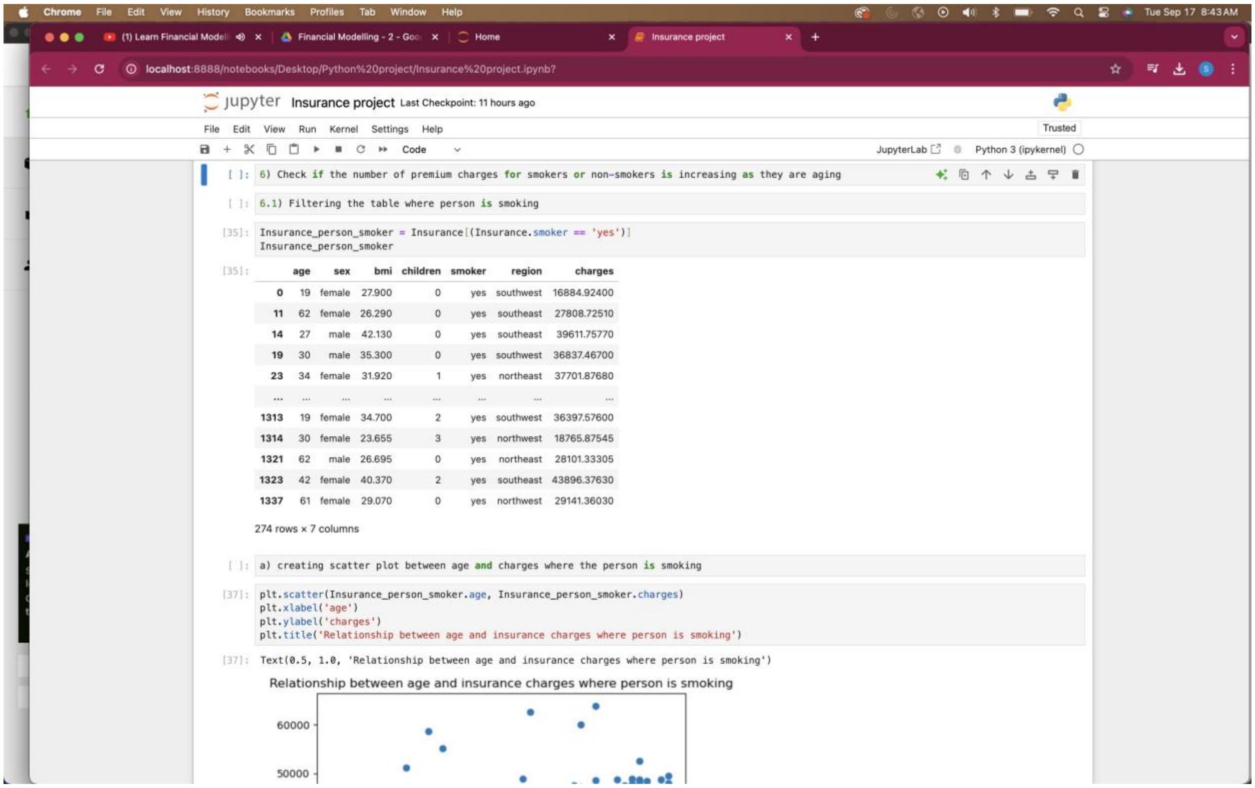
```
[29]: plt.scatter(Insurance.age, Insurance.bmi)
      plt.xlabel('age')
      plt.ylabel('bmi')
      plt.title('Relationship between age and bmi')

[29]: Text(0.5, 1.0, 'Relationship between age and bmi')
```

```
[ ]: b) Categorical variables, creating 2 way cross grid frequency table between sex and region, smoker and region

[31]: pd.crosstab(Insurance.sex,Insurance.region)
```

[31]:

region	northeast	northwest	southeast	southwest
sex				
female	161	164	175	162
male	163	161	189	163

```
[ ]: Observation: The above table provides list of male/female people region wise
```


Chrome File Edit View History Bookmarks Profiles Tab Window Help
Tue Sep 17 8:43 AM
(1) Learn Financial Modell
Financial Modelling - 2 - Goo
Home
Insurance project
localhost:8888/notebooks/Desktop/Python%20project/Insurance%20project.ipynb?
jupyter Insurance project Last Checkpoint: 11 hours ago
File Edit View Run Kernel Settings Help
Trusted
Code
JupyterLab
Python 3 (ipykernel)
Relationship between age and insurance charges where person is smoking
60000
50000

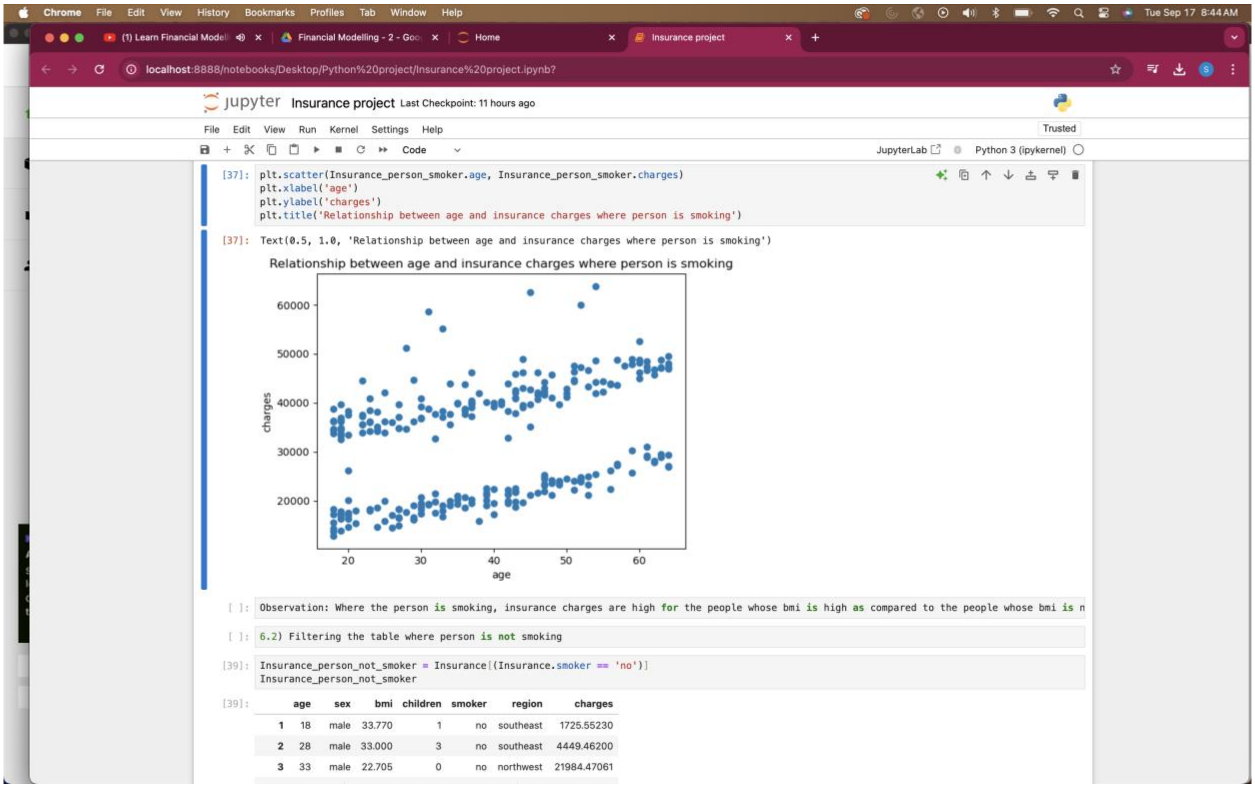
```
[ ]: 6) Check if the number of premium charges for smokers or non-smokers is increasing as they are aging

[ ]: 6.1) Filtering the table where person is smoking

[35]: Insurance_person_smoker = Insurance[(Insurance.smoker == 'yes')]
      Insurance_person_smoker
```

[35]:

	age	sex	bmi	children	smoker	region	charges
0	19	female	27.900	0	yes	southwest	16884.92400
11	62	female	26.290	0	yes	southeast	27808.72510
14	27	male	42.130	0	yes	southeast	39611.75770
19	30	male	35.300	0	yes	southwest	36837.46700
23	34	female	31.920	1	yes	northeast	37701.87680
...	...	...	...	...	...	...	...
1313	19	female	34.700	2	yes	southwest	36397.57600
1314	30	female	23.655	3	yes	northwest	18765.87545
1321	62	male	26.695	0	yes	northeast	28101.33305
1323	42	female	40.370	2	yes	southeast	43896.37630
1337	61	female	29.070	0	yes	northwest	29141.36030

274 rows × 7 columns

```
[ ]: a) creating scatter plot between age and charges where the person is smoking

[37]: plt.scatter(Insurance_person_smoker.age, Insurance_person_smoker.charges)
      plt.xlabel('age')
      plt.ylabel('charges')
      plt.title('Relationship between age and insurance charges where person is smoking')

[37]: Text(0.5, 1.0, 'Relationship between age and insurance charges where person is smoking')
```


Chrome File Edit View History Bookmarks Profiles Tab Window Help
Tue Sep 17 8:44 AM
(1) Learn Financial Modell
Financial Modelling - 2 - Goo
Home
Insurance project
localhost:8888/notebooks/Desktop/Python%20project/Insurance%20project.ipynb?
jupyter Insurance project Last Checkpoint: 11 hours ago
File Edit View Run Kernel Settings Help
Trusted
Code
JupyterLab
Python 3 (ipykernel)
Relationship between age and insurance charges where person is smoking
60000
50000
40000
30000
20000
charges
20
30
40
50
60
age

[37]:
```
plt.scatter(Insurance_person_smoker.age, Insurance_person_smoker.charges)
plt.xlabel('age')
plt.ylabel('charges')
plt.title('Relationship between age and insurance charges where person is smoking')
```

[37]:
```
Text(0.5, 1.0, 'Relationship between age and insurance charges where person is smoking')
```

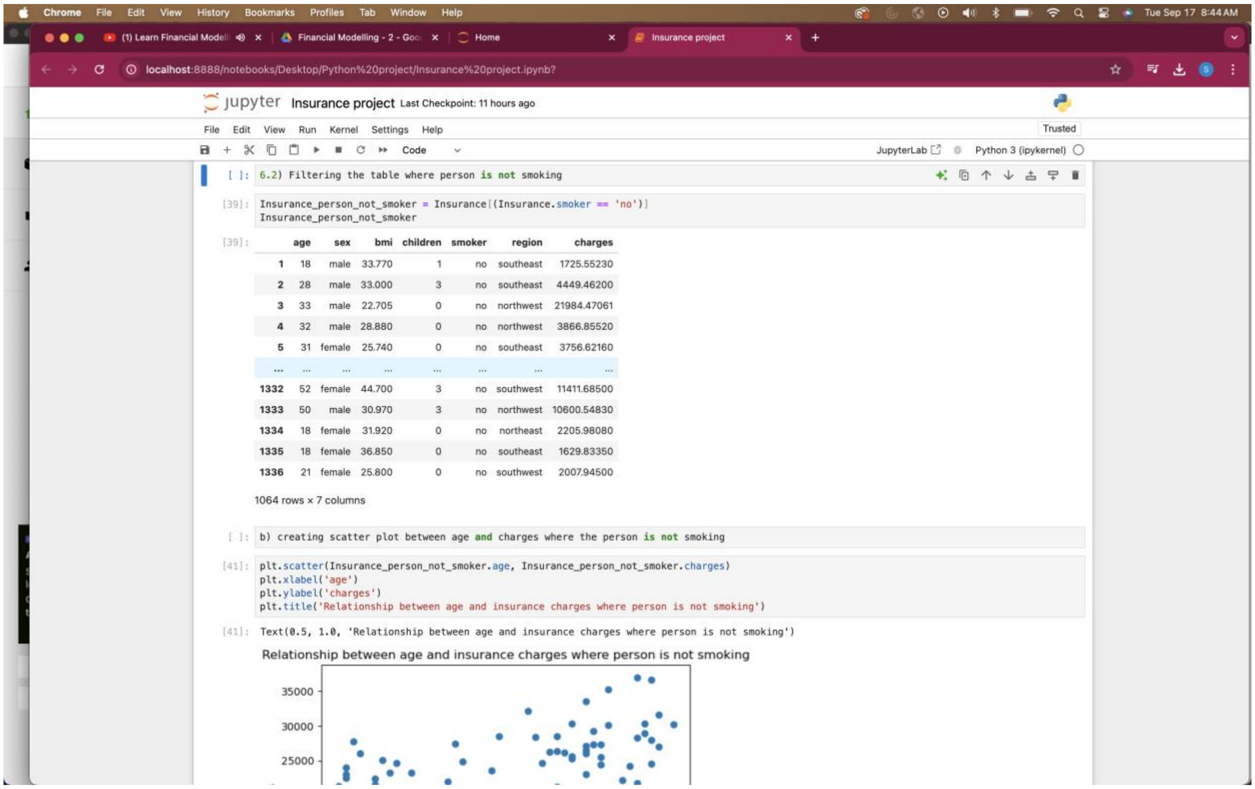
[ ]:
```
Observation: Where the person is smoking, insurance charges are high for the people whose bmi is high as compared to the people whose bmi is n
```

[ ]:
```
6.2) Filtering the table where person is not smoking
```

[39]:
```
Insurance_person_not_smoker = Insurance[(Insurance.smoker == 'no')]
Insurance_person_not_smoker
```

[39]:

	age	sex	bmi	children	smoker	region	charges
1	18	male	33.770	1	no	southeast	1725.55230
2	28	male	33.000	3	no	southeast	4449.46200
3	33	male	22.705	0	no	northwest	21984.47061


Chrome File Edit View History Bookmarks Profiles Tab Window Help
Tue Sep 17 8:44 AM
(1) Learn Financial Modell
Financial Modelling - 2 - Goo
Home
Insurance project
localhost:8888/notebooks/Desktop/Python%20project/Insurance%20project.ipynb?
jupyter Insurance project Last Checkpoint: 11 hours ago
File Edit View Run Kernel Settings Help
Trusted
Code
JupyterLab
Python 3 (ipykernel)
Relationship between age and insurance charges where person is not smoking
35000
30000
25000

[ ]:
```
6.2) Filtering the table where person is not smoking
```

[39]:
```
Insurance_person_not_smoker = Insurance[(Insurance.smoker == 'no')]
Insurance_person_not_smoker
```

[39]:

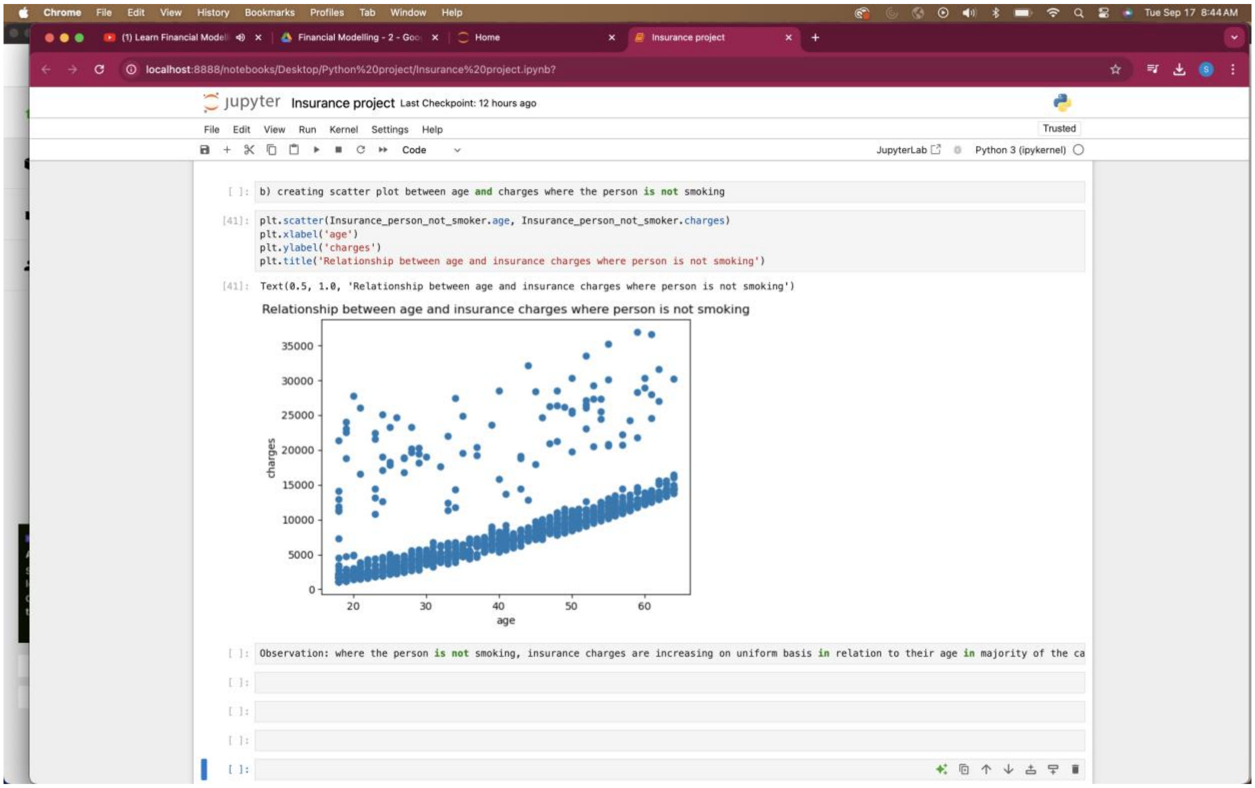
	age	sex	bmi	children	smoker	region	charges
1	18	male	33.770	1	no	southeast	1725.55230
2	28	male	33.000	3	no	southeast	4449.46200
3	33	male	22.705	0	no	northwest	21984.47061
4	32	male	28.880	0	no	northwest	3866.85520
5	31	female	25.740	0	no	southeast	3756.62160
...	...	...	...	...	...	...	...
1332	52	female	44.700	3	no	southwest	11411.68500
1333	50	male	30.970	3	no	northwest	10600.54830
1334	18	female	31.920	0	no	northeast	2205.98080
1335	18	female	36.850	0	no	southeast	1629.83350
1336	21	female	25.800	0	no	southwest	2007.94500

1064 rows × 7 columns

[ ]:
```
b) creating scatter plot between age and charges where the person is not smoking
```

[41]:
```
plt.scatter(Insurance_person_not_smoker.age, Insurance_person_not_smoker.charges)
plt.xlabel('age')
plt.ylabel('charges')
plt.title('Relationship between age and insurance charges where person is not smoking')
```

[41]:
```
Text(0.5, 1.0, 'Relationship between age and insurance charges where person is not smoking')
```



Chrome File Edit View History Bookmarks Profiles Tab Window Help
Tue Sep 17 8:44 AM
(1) Learn Financial Modell
Financial Modelling - 2 - Goo
Home
Insurance project
localhost:8888/notebooks/Desktop/Python%20project/Insurance%20project.ipynb?
jupyter Insurance project Last Checkpoint: 12 hours ago
File Edit View Run Kernel Settings Help
Trusted
Code
JupyterLab
Python 3 (ipykernel)

```
[ ]: b) creating scatter plot between age and charges where the person is not smoking
```

```
[41]: plt.scatter(Insurance_person_not_smoker.age, Insurance_person_not_smoker.charges)
      plt.xlabel('age')
      plt.ylabel('charges')
      plt.title('Relationship between age and insurance charges where person is not smoking')
```

```
[41]: Text(0.5, 1.0, 'Relationship between age and insurance charges where person is not smoking')
```

Relationship between age and insurance charges where person is not smoking
35000
30000
25000
20000
15000
10000
5000
0
charges
20
30
40
50
60
age

```
[ ]: Observation: where the person is not smoking, insurance charges are increasing on uniform basis in relation to their age in majority of the ca
```

```
[ ]:
[ ]:
[ ]:
[ ]:
```