

```
In [1]: '''
CS 418: Final Project
Title - US Census Demographic Data
Authors: Anusha Sagi, Fatima Kahack, Lydia Tse
Description: The following is code to analyze the US Census Data and preprocess it.
'''
```

```
Out[1]: '\nCS 418: Final Project\nTitle - US Census Demographic Data\nAuthors: Anusha Sagi, Fatima Kahack, Lydia Tse\nDescription: The following is code to analyze the US Census Data and preprocess it.\n'
```

```
In [2]: # Load libraries
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
```

```
In [3]: # Load US Census Demographic Data
census_data = pd.read_csv('US_census_data_2017.csv')
census_data.head()
```

```
Out[3]:
```

	CountyId	State	County	TotalPop	Men	Women	Hispanic	White	Black	Native	...	Walk	OtherTransp	WorkAtHome	MeanCommute	Employed	Private
0	1001	Alabama	Autauga County	55036	26899	28137	2.7	75.4	18.9	0.3	...	0.6	1.3	2.5	25.8	24112	
1	1003	Alabama	Baldwin County	203360	99527	103833	4.4	83.1	9.5	0.8	...	0.8	1.1	5.6	27.0	89527	
2	1005	Alabama	Barbour County	26201	13976	12225	4.2	45.7	47.8	0.2	...	2.2	1.7	1.3	23.4	8878	
3	1007	Alabama	Bibb County	22580	12251	10329	2.4	74.6	22.0	0.4	...	0.3	1.7	1.5	30.0	8171	
4	1009	Alabama	Blount County	57667	28490	29177	9.0	87.4	1.5	0.3	...	0.4	0.4	2.1	35.0	21380	

5 rows × 37 columns

```
In [4]: #Divide into train and test datasets
msk = np.random.rand(len(census_data)) <= 0.75
train = census_data[msk]
test = census_data[~msk]
```

```
In [5]: #Determine Length of train and test datasets
print(len(test))
print(len(train))
```

778  
2442

```
In [6]: census_train = pd.DataFrame(data=train)
census_train.to_csv("census_train.csv")
census_train.head()
```

```
Out[6]:
```

	CountyId	State	County	TotalPop	Men	Women	Hispanic	White	Black	Native	...	Walk	OtherTransp	WorkAtHome	MeanCommute	Employed	Private
1	1003	Alabama	Baldwin County	203360	99527	103833	4.4	83.1	9.5	0.8	...	0.8	1.1	5.6	27.0	89527	
3	1007	Alabama	Bibb County	22580	12251	10329	2.4	74.6	22.0	0.4	...	0.3	1.7	1.5	30.0	8171	
6	1013	Alabama	Butler County	20126	9416	10710	0.3	52.2	44.7	0.1	...	0.9	0.9	2.0	23.2	7727	
7	1015	Alabama	Calhoun County	115527	55593	59934	3.6	72.7	20.4	0.2	...	1.3	1.1	3.2	24.8	47392	
9	1019	Alabama	Cherokee County	25855	12862	12993	1.6	91.8	5.0	0.5	...	0.3	0.3	2.0	26.5	9879	

5 rows × 37 columns

```
In [7]: census_test = pd.DataFrame(data=test)
census_test.to_csv("census_test.csv")
census_test.head()
```

Out[7]:

	CountyId	State	County	TotalPop	Men	Women	Hispanic	White	Black	Native	...	Walk	OtherTransp	WorkAtHome	MeanCommute	Employed	Priv
0	1001	Alabama	Autauga County	55036	26899	28137	2.7	75.4	18.9	0.3	...	0.6	1.3	2.5	25.8	24112	
2	1005	Alabama	Barbour County	26201	13976	12225	4.2	45.7	47.8	0.2	...	2.2	1.7	1.3	23.4	8878	
4	1009	Alabama	Blount County	57667	28490	29177	9.0	87.4	1.5	0.3	...	0.4	0.4	2.1	35.0	21380	
5	1011	Alabama	Bullock County	10478	5616	4862	0.3	21.6	75.6	1.0	...	6.2	1.7	3.0	29.8	4290	
8	1017	Alabama	Chambers County	33895	16320	17575	2.2	56.2	39.3	0.3	...	0.6	0.5	2.0	23.6	14527	

5 rows x 37 columns



```
In [ ]:
```

```
In [ ]:
```