

Assignment Week 6

This week we are using Pandas.

In [10]:

```
%matplotlib inline
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
plt.style.use('ggplot')
```

RMS Titanic

One file is provided:

- `titanic.csv` - this file contains titanic passenger information

The file `titanic.csv` provides passenger information on the ship RMS Titanic that sank on her maiden voyage on April 15, 1912.

1. Read the passenger information of the Titanic. Show an overview of *unique* last names of the passenger list (so ignore the first names and the "Mr", "Mrs.", etc) and determine the number of different last names.

In [11]:

```
df = pd.read_csv('titanic.csv')
df['Last_Name'] = '' # Add a new column called Last_Name in dataframe

for i in range(len(df)):
    name_split = df.loc[i, 'Name'].split(',', 1) # Split the name at the comma
    l_name = name_split[0] # First element is the last name
    df.loc[i, 'Last_Name'] = l_name # Add to the Last_Name column

unique_last_name = df['Last_Name'].unique() # Keep only unique last names in an array
print(unique_last_name[:10]) # Show the first 10 unique last names
print('The number of different last names is:', len(unique_last_name))
```

```
['Kelly' 'Wilkes' 'Myles' 'Wirz' 'Hirvonen' 'Svensson' 'Connolly'
 'Caldwell' 'Abraham' 'Davies']
The number of different last names is: 352
```

2. Assuming that passengers with the same last name form a family, determine how big the largest family aboard the Titanic was.

In [12]:

```
max_number = max(df['Last_Name'].value_counts()) # Count the frequency of each last name,
# then find the highest frequency
print('The largest family onboard has', max_number, 'members.')
```

The largest family onboard has 4 members.

3. Drop all the rows for which the Fare is 0 or NaN. Store the result.

In [13]:

```
df2 = df.dropna()
df3 = df2[df2.Fare != 0]

print(df3)
```

	PassengerId	Pclass	Name \
12	904	1	Snyder, Mrs. John Pillsbury (Nelle Stevenson)
14	906	1	Chaffee, Mrs. Herbert Fuller (Carrie Constance...
24	916	1	Ryerson, Mrs. Arthur Larned (Emily Maria Borie)
26	918	1	Ostby, Miss. Helene Ragnhild
28	920	1	Brady, Mr. John Bertram
34	926	1	Mock, Mr. Philipp Edmund
44	936	1	Kimball, Mrs. Edwin Nelson Jr (Gertrude Parsons)
46	938	1	Chevre, Mr. Paul Romaine
48	940	1	Bucknell, Mrs. William Robert (Emma Eliza Ward)
50	942	1	Smith, Mr. Lucien Philip
53	945	1	Fortune, Miss. Ethel Flora
57	949	3	Abelseth, Mr. Olaus Jorgensen
59	951	1	Chaudanson, Miss. Victorine
64	956	1	Ryerson, Master. John Borie
68	960	1	Tucker, Mr. Gilbert Milligan Jr
69	961	1	Fortune, Mrs. Mark (Mary McDougald)
73	965	1	Ovies y Rodriguez, Mr. Servando
74	966	1	Geiger, Miss. Amalie
75	967	1	Keeping, Mr. Edwin
77	969	1	Cornell, Mrs. Robert Clifford (Malvina Helen L...
81	973	1	Straus, Mr. Isidor
92	984	1	Davidson, Mrs. Thornton (Orian Hays)
96	988	1	Cavendish, Mrs. Tyrell William (Julia Florence...
100	992	1	Stengel, Mrs. Charles Emil Henry (Annie May Mo...
109	1001	2	Swane, Mr. George
112	1004	1	Evans, Miss. Edith Corse
114	1006	1	Straus, Mrs. Isidor (Rosalie Ida Blun)
117	1009	3	Sandstrom, Miss. Beatrice Irene
118	1010	1	Beattie, Mr. Thomson
122	1014	1	Schabert, Mrs. Paul (Emma Mock)
..
287	1179	1	Snyder, Mr. John Pillsbury
293	1185	1	Dodge, Dr. Washington
305	1197	1	Crosby, Mrs. Edward Gifford (Catherine Elizabe...
306	1198	1	Allison, Mr. Hudson Joshua Creighton
308	1200	1	Hays, Mr. Charles Melville
314	1206	1	White, Mrs. John Stuart (Ella Holmes)
316	1208	1	Spencer, Mr. William Augustus
321	1213	3	Krekorian, Mr. Neshan
322	1214	2	Nesson, Mr. Israel
326	1218	2	Becker, Miss. Ruth Elizabeth
331	1223	1	Dulles, Mr. William Crothers
335	1227	1	Maguire, Mr. John Edward
343	1235	1	Cardeza, Mrs. James Warburton Martinez (Charlo...
350	1242	1	Greenfield, Mrs. Leo David (Blanche Strouse)
355	1247	1	Julian, Mr. Henry Forbes
356	1248	1	Brown, Mrs. John Murray (Caroline Lane Lamson)
364	1256	1	Harder, Mrs. George Achilles (Dorothy Annan)
371	1263	1	Wilson, Miss. Helen Alice
374	1266	1	Dodge, Mrs. Washington (Ruth Vidaver)
378	1270	1	Hinkins. Mr. William Edward

390	1282	1	Payne, Mr. Vivian Ponsonby
391	1283	1	Lines, Mrs. Ernest H (Elizabeth Lindsey James)
395	1287	1	Smith, Mrs. Lucien Philip (Mary Eloise Hughes)
397	1289	1	Frolicher-Stehli, Mrs. Maxmillian (Margaretha ...
400	1292	1	Bonnell, Miss. Caroline
404	1296	1	Frauenthal, Mr. Isaac Gerald
405	1297	2	Nourney, Mr. Alfred (Baron von Drachstedt)"
407	1299	1	Widener, Mr. George Dunton
411	1303	1	Minahan, Mrs. William Edward (Lillian E Thorpe)
414	1306	1	Oliva y Ocana, Dona. Fermina

	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin \
12	female	23.0	1	0	21228	82.2667	B45
14	female	47.0	1	0	W.E.P. 5734	61.1750	E31
24	female	48.0	1	3	PC 17608	262.3750	B57 B59 B63 B66
26	female	22.0	0	1	113509	61.9792	B36
28	male	41.0	0	0	113054	30.5000	A21
34	male	30.0	1	0	13236	57.7500	C78
44	female	45.0	1	0	11753	52.5542	D19
46	male	45.0	0	0	PC 17594	29.7000	A9
48	female	60.0	0	0	11813	76.2917	D15
50	male	24.0	1	0	13695	60.0000	C31
53	female	28.0	3	2	19950	263.0000	C23 C25 C27
57	male	25.0	0	0	348122	7.6500	F G63
59	female	36.0	0	0	PC 17608	262.3750	B61
64	male	13.0	2	2	PC 17608	262.3750	B57 B59 B63 B66
68	male	31.0	0	0	2543	28.5375	C53
69	female	60.0	1	4	19950	263.0000	C23 C25 C27
73	male	28.5	0	0	PC 17562	27.7208	D43
74	female	35.0	0	0	113503	211.5000	C130
75	male	32.5	0	0	113503	211.5000	C132
77	female	55.0	2	0	11770	25.7000	C101
81	male	67.0	1	0	PC 17483	221.7792	C55 C57
92	female	27.0	1	2	F.C. 12750	52.0000	B71
96	female	76.0	1	0	19877	78.8500	C46
100	female	43.0	1	0	11778	55.4417	C116
109	male	18.5	0	0	248734	13.0000	F
112	female	36.0	0	0	PC 17531	31.6792	A29
114	female	63.0	1	0	PC 17483	221.7792	C55 C57
117	female	1.0	1	1	PP 9549	16.7000	G6
118	male	36.0	0	0	13050	75.2417	C6
122	female	35.0	1	0	13236	57.7500	C28
..
287	male	24.0	1	0	21228	82.2667	B45
293	male	53.0	1	1	33638	81.8583	A34
305	female	64.0	1	1	112901	26.5500	B26
306	male	30.0	1	2	113781	151.5500	C22 C26
308	male	55.0	1	1	12749	93.5000	B69
314	female	55.0	0	0	PC 17760	135.6333	C32
316	male	57.0	1	0	PC 17569	146.5208	B78
321	male	25.0	0	0	2654	7.2292	F E57
322	male	26.0	0	0	244368	13.0000	F2
326	female	12.0	2	1	230136	39.0000	F4
331	male	39.0	0	0	PC 17580	29.7000	A18
335	male	30.0	0	0	110469	26.0000	C106
343	female	58.0	0	1	PC 17755	512.3292	B51 B53 B55
350	female	45.0	0	1	PC 17759	63.3583	D10 D12
355	male	50.0	0	0	113044	26.0000	E60
356	female	59.0	2	0	11769	51.4792	C101
364	female	25.0	1	0	11765	55.4417	E50
371	female	31.0	0	0	16966	134.5000	E39 E41
374	female	54.0	1	1	33638	81.8583	A34
378	male	55.0	0	0	680	50.0000	C39

390	male	23.0	0	0		12749	93.5000	B24
391	female	51.0	0	1	PC	17592	39.4000	D28
395	female	18.0	1	0		13695	60.0000	C31
397	female	48.0	1	1		13567	79.2000	B41
400	female	30.0	0	0		36928	164.8667	C7
404	male	43.0	1	0		17765	27.7208	D40
405	male	20.0	0	0	SC/PARIS	2166	13.8625	D38
407	male	50.0	1	1		113503	211.5000	C80
411	female	37.0	1	0		19928	90.0000	C78
414	female	39.0	0	0	PC	17758	108.9000	C105

	Embarked	Last_Name
12	S	Snyder
14	S	Chaffee
24	C	Ryerson
26	C	Ostby
28	S	Brady
34	C	Mock
44	S	Kimball
46	C	Chevre
48	C	Bucknell
50	S	Smith
53	S	Fortune
57	S	Abelseth
59	C	Chaudanson
64	C	Ryerson
68	C	Tucker
69	S	Fortune
73	C	Ovies y Rodriguez
74	C	Geiger
75	C	Keeping
77	S	Cornell
81	S	Straus
92	S	Davidson
96	S	Cavendish
100	C	Stengel
109	S	Swane
112	C	Evans
114	S	Straus
117	S	Sandstrom
118	C	Beattie
122	C	Schabert
..
287	S	Snyder
293	S	Dodge
305	S	Crosby
306	S	Allison
308	S	Hays
314	C	White
316	C	Spencer
321	C	Krekorian
322	S	Nesson
326	S	Becker
331	C	Dulles
335	S	Maguire
343	C	Cardeza
350	C	Greenfield
355	S	Julian
356	S	Brown
364	C	Harder
371	C	Wilson
374	S	Dodge
378	S	Hipkins
390	S	Pavne

```

391      S      Lines
395      S      Smith
397      C  Frolicher-Stehli
400      S      Bonnell
404      C      Frauenthal
405      C      Nourney
407      C      Widener
411      Q      Minahan
414      C      Oliva y Ocana

```

[86 rows x 12 columns]

4. Show the average Fare price per Passenger Class (Pclass) in a bar chart.

In [14]:

```

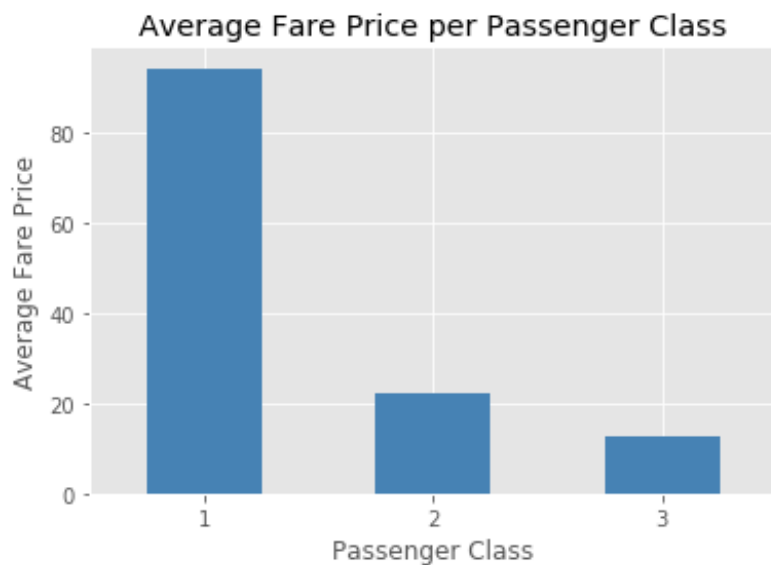
df4 = df.groupby('Pclass')['Fare'].mean()
plot1 = df4.plot.bar(title = "Average Fare Price per Passenger Class", rot = 0, color = '#4682B4')

plot1.set_xlabel("Passenger Class")
plot1.set_ylabel("Average Fare Price")

```

Out[14]:

Text(0, 0.5, 'Average Fare Price')



5. Plot a histogram of the passenger's age, grouped by sex. So the end-result should show separate histograms for males and females.

In [27]:

```

plot2 = df['Age'].hist(by=df['Sex'], rot = 0)
plt.suptitle("Histogram of the passenger's age, grouped by sex", y = 1.1, size = 16)

```

Out[27]:

Text(0.5, 1.1, "Histogram of the passenger's age, grouped by sex")

Histogram of the passenger's age, grouped by sex

