The Titanic Project

The Titanic was a luxurious British liner that sunk in the North Atlantic Ocean on April 15, 1912, after colliding with an iceberg during its first voyage from Southampton, UK to New York City, USA. It was a

marvel of modern engineering and the largest passenger ship at the time. However, the disaster resulted in the loss of over 1,500 lives due to a shortage of lifeboats and inadequate safety procedures. This made it one of the deadliest maritime disasters in history. Today, the story of the Titanic lives on and has been retold in numerous books, movies, and documentaries.

THE DATASET

The Titanic dataset is a well-known dataset used for analyzing data related to the passengers onboard the ship, including their demographics, ticket information, cabin details, and survival status.

The dataset for this study was obtained from Kaggle and you can download it HERE. The data contains the following:

- **Survival** (0=No, 1=Yes)
- **Passenger Class** (**Pclass**) [1=1st Class, 2=2nd Class, 3=3rd Class]
- **Sex** = Sex of Passenger
- **Age** = Age of passenger in years
- **SibSp** = Number of siblings and spouses onboard the Titanic
- **Parch** = Number of parents and children aboard the Titanic
- **ticket** = ticket number
- **Cabin** = Cabin number
- **Embarked** = Point of embarkation [C = Cherbourg, Q = Queenstown, S = Southampton]

So, please continue reading as I delve into the fascinating world of the Titanic dataset analysis.

PROBLEM STATEMENT

The purpose of this analysis is to predict the likelihood of survival for Titanic passengers based on various factors such as age, gender, ticket number, and other variables. The dataset includes information about 1309 passengers, which is a subset of the total 2,224 individuals that comprised the passengers and crew members of the Titanic. This problem statement is significant as it provides insight into the variables that influenced the survival rates of the passengers during the tragic incident.

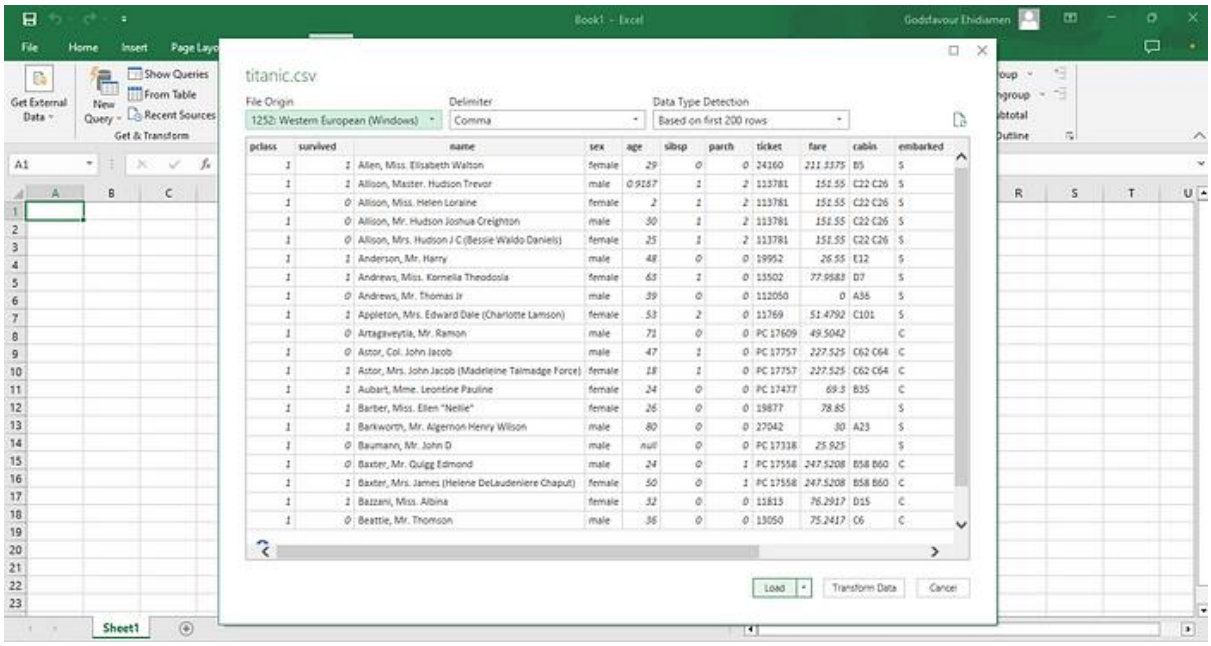
The analysis involved importing and cleaning data, predicting passenger survival, and presenting conclusions.

Data Exploration Tool

I used ***Microsoft Excel*** for my data exploration and analysis.

STEP 1: DATA IMPORTATION

Importing the dataset into your worksheet is the initial step in data analytics. The dataset I needed was available in CSV format as a "zip file". I downloaded the file to my computer after unzipping it on my Google Drive and opened it in Excel. To do so, I clicked on the "File" menu and selected "Open".



Loading the data

The display above popped up; then I clicked ***Load***. The original data set is displayed below.

	pclass	survived	name	sex	age	sibsp	parch	ticket	fare	cabin	embarked	boat	body	home.dest
2	1	1	Allen, Miss. Elisabeth Walton	female	29	0	0	24160	211.3375	B5	S	2		St Louis, MO
3	1	1	Allison, Master. Hudson Trevor	male	0.9167	1	2	113781	151.55	C22 C26	S	11		Montreal, PQ / Chesterville, ON
4	1	0	Allison, Miss. Helen Loraine	female	2	1	2	113781	151.55	C22 C26	S			Montreal, PQ / Chesterville, ON
5	1	0	Allison, Mr. Hudson Joshua Creighton	male	30	1	2	113781	151.55	C22 C26	S		135	Montreal, PQ / Chesterville, ON
6	1	0	Allison, Mrs. Hudson J C (Bessie Waldo Daniels)	female	25	1	2	113781	151.55	C22 C26	S			Montreal, PQ / Chesterville, ON
7	1	1	Anderson, Mr. Harry	male	48	0	0	19952	26.55	E12	S	3		New York, NY
8	1	1	Andrews, Miss. Kornelia Theodosia	female	63	1	0	13502	77.9583	D7	S	10		Hudson, NY
9	1	0	Andrews, Mr. Thomas Jr	male	39	0	0	112050	0	A36	S			Belfast, NI
10	1	1	Appleton, Mrs. Edward Dale (Charlotte Lamson)	female	53	2	0	11769	51.4792	C101	S	D		Bayside, Queens, NY
11	1	0	Artagaveytia, Mr. Ramon	male	71	0	0	PC 17609	49.5042		C		22	Montevideo, Uruguay
12	1	0	Astor, Col. John Jacob	male	47	1	0	PC 17757	227.525	C62 C64	C		124	New York, NY
13	1	1	Astor, Mrs. John Jacob (Madeleine Talmadge Force)	female	18	1	0	PC 17757	227.525	C62 C64	C	4		New York, NY
14	1	1	Aubart, Mme. Leontine Pauline	female	24	0	0	PC 17477	69.3	B35	C	9		Paris, France
15	1	1	Barber, Miss. Ellen "Nellie"	female	26	0	0	19877	78.85		S	6		
16	1	1	Barkworth, Mr. Algernon Henry Wilson	male	80	0	0	27042	30	A23	S	B		Hessle, Yorks
17	1	0	Baumann, Mr. John D	male		0	0	PC 17318	25.925		S			New York, NY
18	1	0	Baxter, Mr. Quigg Edmond	male	24	0	1	PC 17558	247.5208	B58 B60	C			Montreal, PQ
19	1	1	Baxter, Mrs. James (Helene DeLaudeniere Chaput)	female	50	0	1	PC 17558	247.5208	B58 B60	C	6		Montreal, PQ
20	1	1	Bazzani, Miss. Albina	female	32	0	0	11813	76.2917	D15	C	8		

The raw data

STEP 2: DATA CLEANING

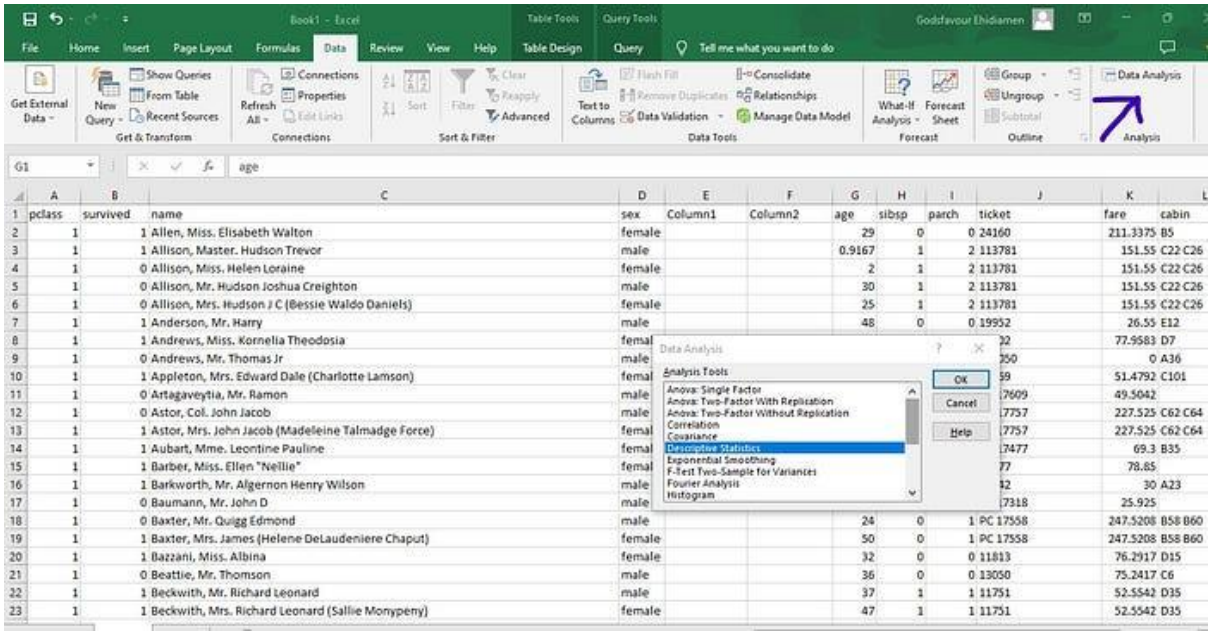
During the examination, I noticed missing values in the age column. Missing values can be problematic for data analysis. Hence, it is important to identify and handle them appropriately.

Age Column

To handle the issue of missing values in the age column of the dataset, I decided to replace them with the average age of all the passengers on board. In order to facilitate my statistical analysis, I created two helper columns to the left of the age column. To achieve this, I simply selected the age column, right-clicked, and chose the "insert" option twice. These new columns allowed me to carry out my analysis without any difficulty.

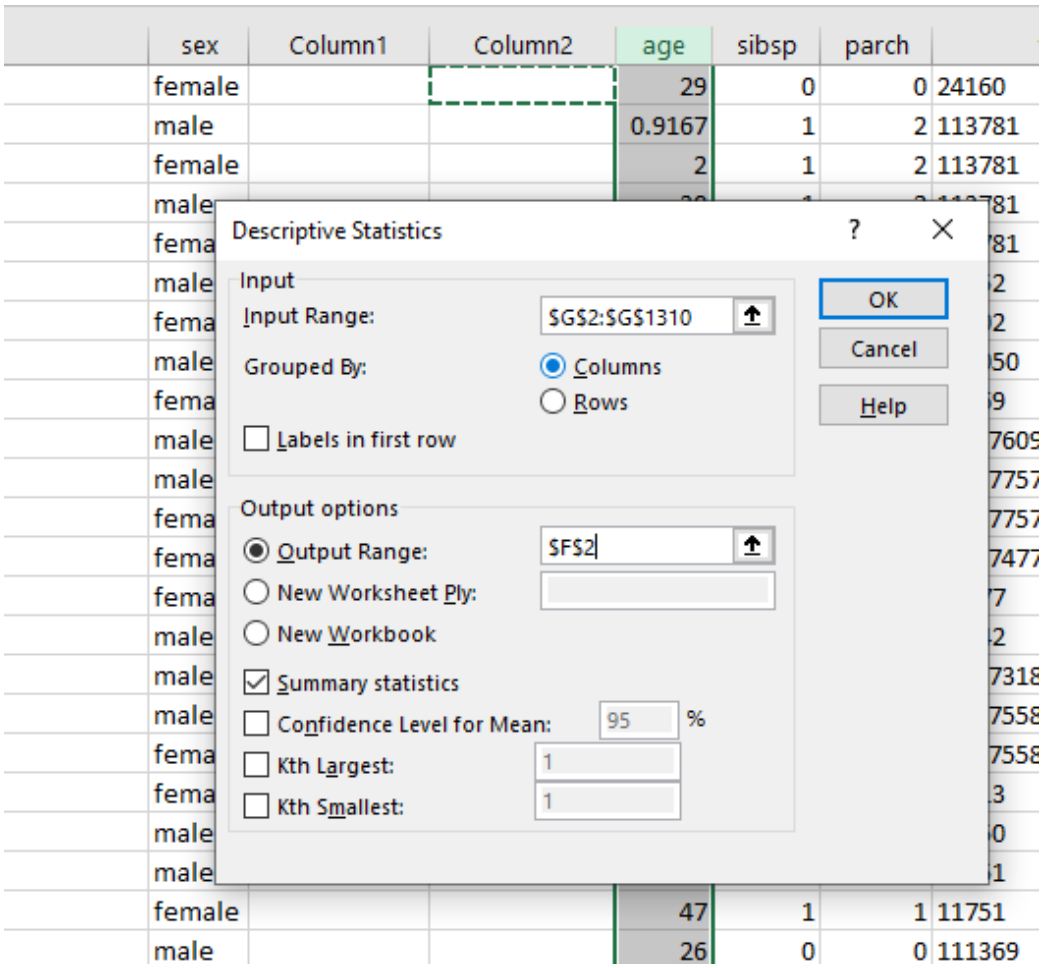
D	E	F	G	
sex	Column1	Column2	age	si
female			29	
male			0.9167	
female			2	
male			30	
female			25	
male			48	
female			63	
male			39	
female			53	
male			71	
male			47	
female			18	
female			24	
female			26	
male			80	
male				
male			24	
female			50	
female			32	
male			36	
male			37	
female			47	

Then, I carried out ***descriptive statistics***. I first selected the age data and clicked on data analysis found in the data ribbon (as seen below). Then select 'OK.'



The data analytics function

It showed a drop-down where I filled in the input and output ranges. After this, I clicked OK.

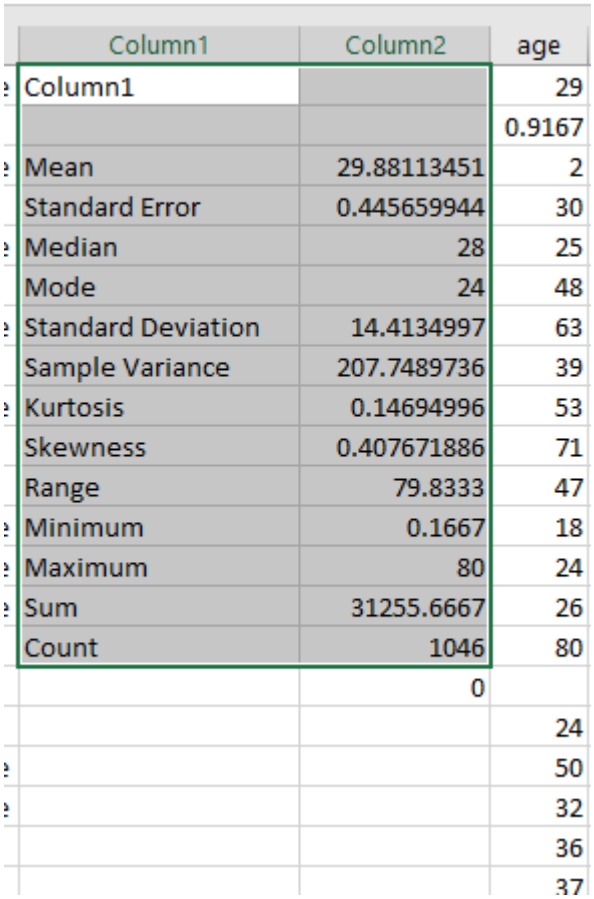


Descriptive analysis

When working with Excel, the input range refers to the range of cells that contains the data you want to analyze or perform calculations on. On the other hand, the output range refers to the range of cells where the results of a formula, function, or data analysis tool will be displayed.

It is crucial to ensure that the output range is large enough to accommodate all the results of the analysis tool. If the output range is insufficient, you may encounter an error message, which can hinder your work. Therefore, it's always best to double-check the output range's size before applying any formula or function to avoid such errors.

See result below.

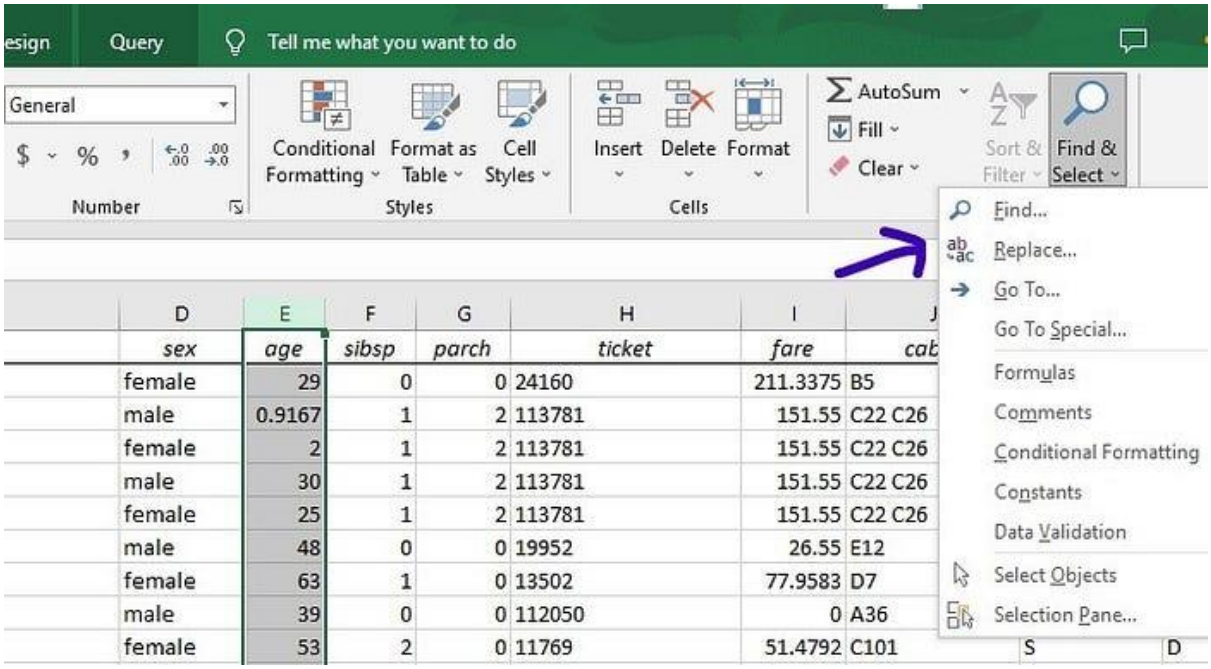


Column1	Column2	age
Column1		29
		0.9167
Mean	29.88113451	2
Standard Error	0.445659944	30
Median	28	25
Mode	24	48
Standard Deviation	14.4134997	63
Sample Variance	207.7489736	39
Kurtosis	0.14694996	53
Skewness	0.407671886	71
Range	79.8333	47
Minimum	0.1667	18
Maximum	80	24
Sum	31255.6667	26
Count	1046	80
	0	
		24
		50
		32
		36
		37

Based on the information provided in the image above, the average age is 28.88.

Find & Select

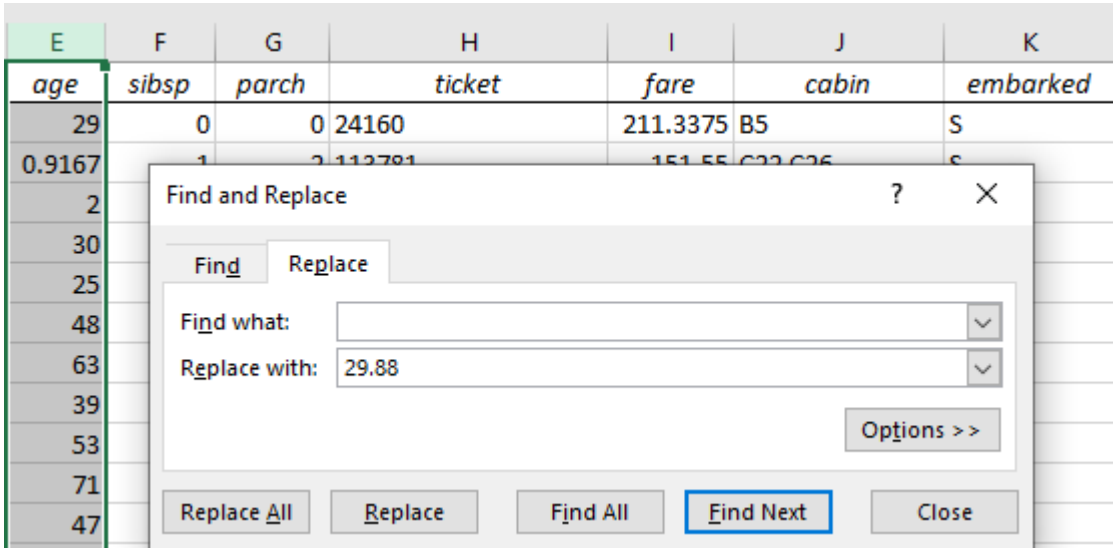
Afterward, I removed the two helper columns and filled in the missing ages using the "Find and Select" option. To accomplish this, select the age column, go to the "Home" tab, find the "Find & Select" dropdown menu, and click on "Replace". As shown in the image below.



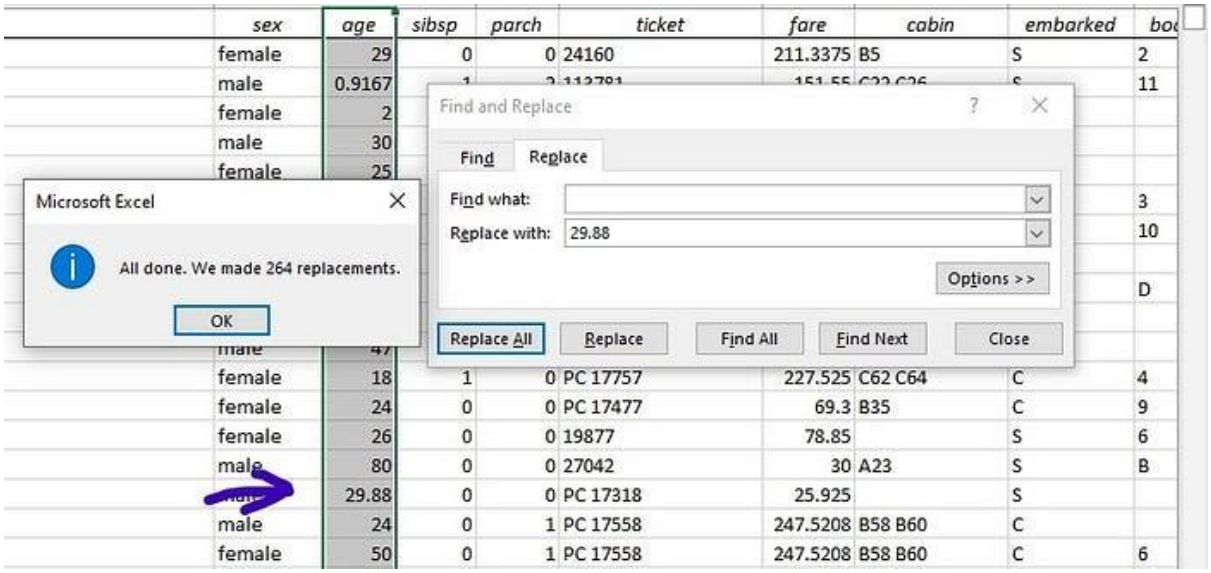
	D	E	F	G	H	I	J
	sex	*age*	*sibsp*	*parch*	*ticket*	*fare*	*cab*
	female	29	0	0	24160	211.3375	B5
	male	0.9167	1	2	113781	151.55	C22 C26
	female	2	1	2	113781	151.55	C22 C26
	male	30	1	2	113781	151.55	C22 C26
	female	25	1	2	113781	151.55	C22 C26
	male	48	0	0	19952	26.55	E12
	female	63	1	0	13502	77.9583	D7
	male	39	0	0	112050	0	A36
	female	53	2	0	11769	51.4792	C101

S D

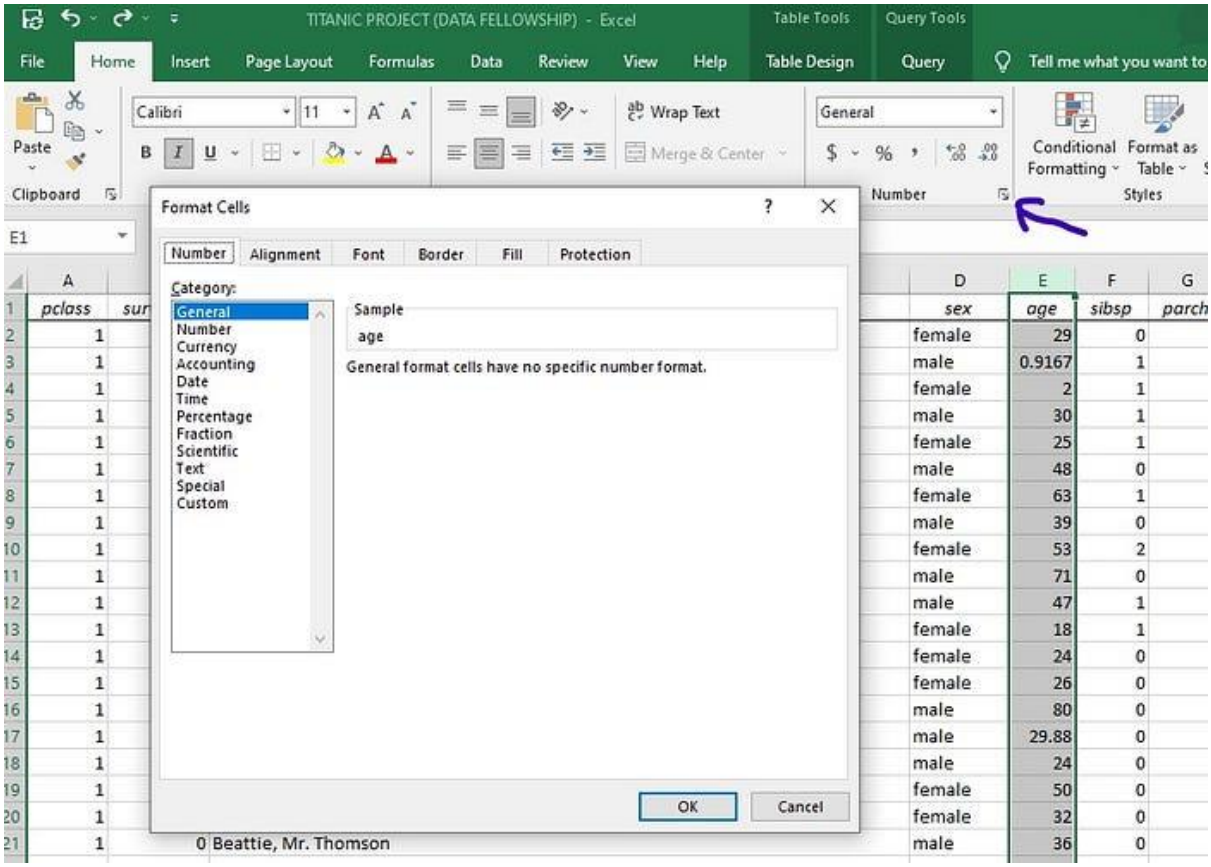
This action resulted in the image below. I left the "Find What" field blank, but filled the "Replace" field with the mean figure (29.88). Then, I selected "Replace All".



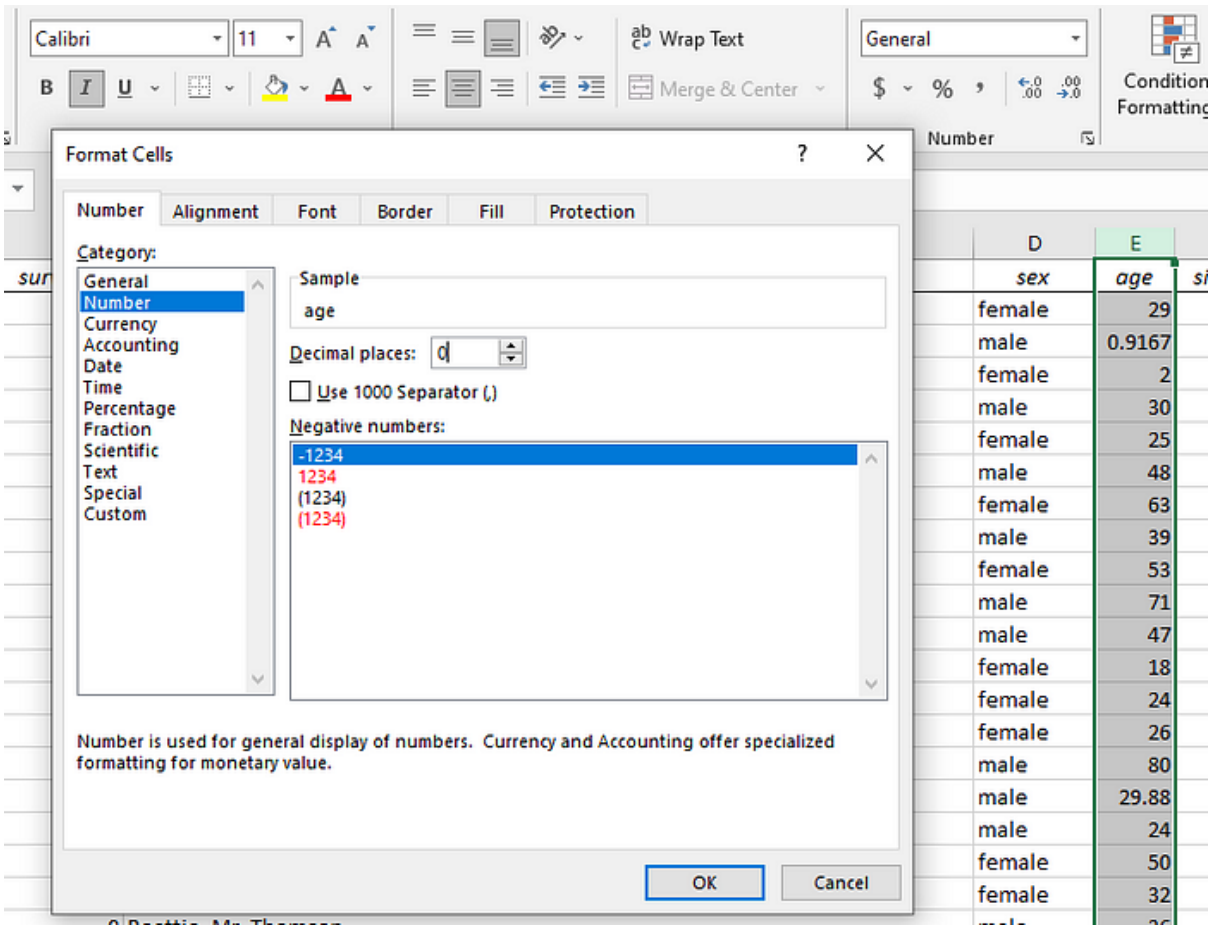
Find and replace function



From the image displayed above, it is evident that the "Find & Select" option has replaced all the empty cells with the value 29.88. After this, click on the "OK" button and exit the drop-down. However, it is not a common practice to have a person's age in decimal points as age is a whole number. Therefore, I rounded off the ages in decimals to the nearest whole number. For instance, if someone is 25.7 years old, the age in a whole number will be 26. Once I became aware of this, I selected the age column, clicked on the "Home" tab, and then clicked the little arrow located near the "Number Format" option. As shown in the image below,



After that, I clicked on "Number" and changed the decimal places to 0. As shown below. Then I clicked "OK".



Changing to the nearest whole number

Hurray! The ages are now in their whole numbers. *See image below*

sex	age	sib
female	29	
male	1	
female	2	
male	30	
female	25	
male	48	
female	63	
male	39	
female	53	
male	71	
male	47	
female	18	
female	24	
female	26	
male	80	
male	30	
male	24	
female	50	
female	32	
male	36	
male	37	
female	47	

After some analysis, I noticed that it would be helpful to categorize the ages of the passengers. Therefore, I added a new column to the dataset and grouped the ages into three categories. For example, individuals with ages ranging from 0 to 18 were classified as children, while those between 18 to 35 years were classified as young adults. Additionally, passengers aged 35 years and above were grouped as adults.

```
=IF(E2<18,"Child",IF(E2<35,"Young Adult","Adult"))
```

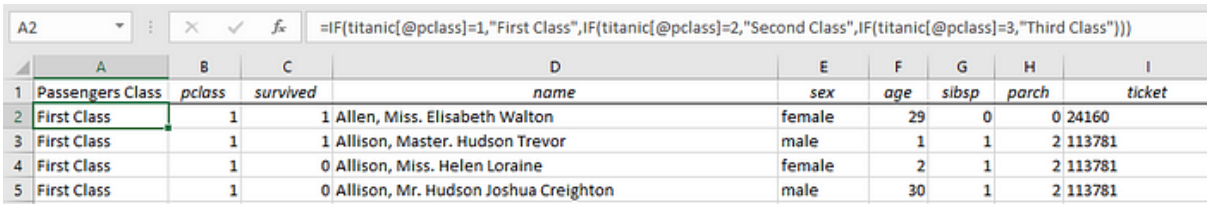
Following that, I visited the **P Class** column.

P Class Column

In the dataset, the class of passengers was originally represented by numerical values 1, 2, and 3. However, to make the data more understandable, I converted these numerical values to textual

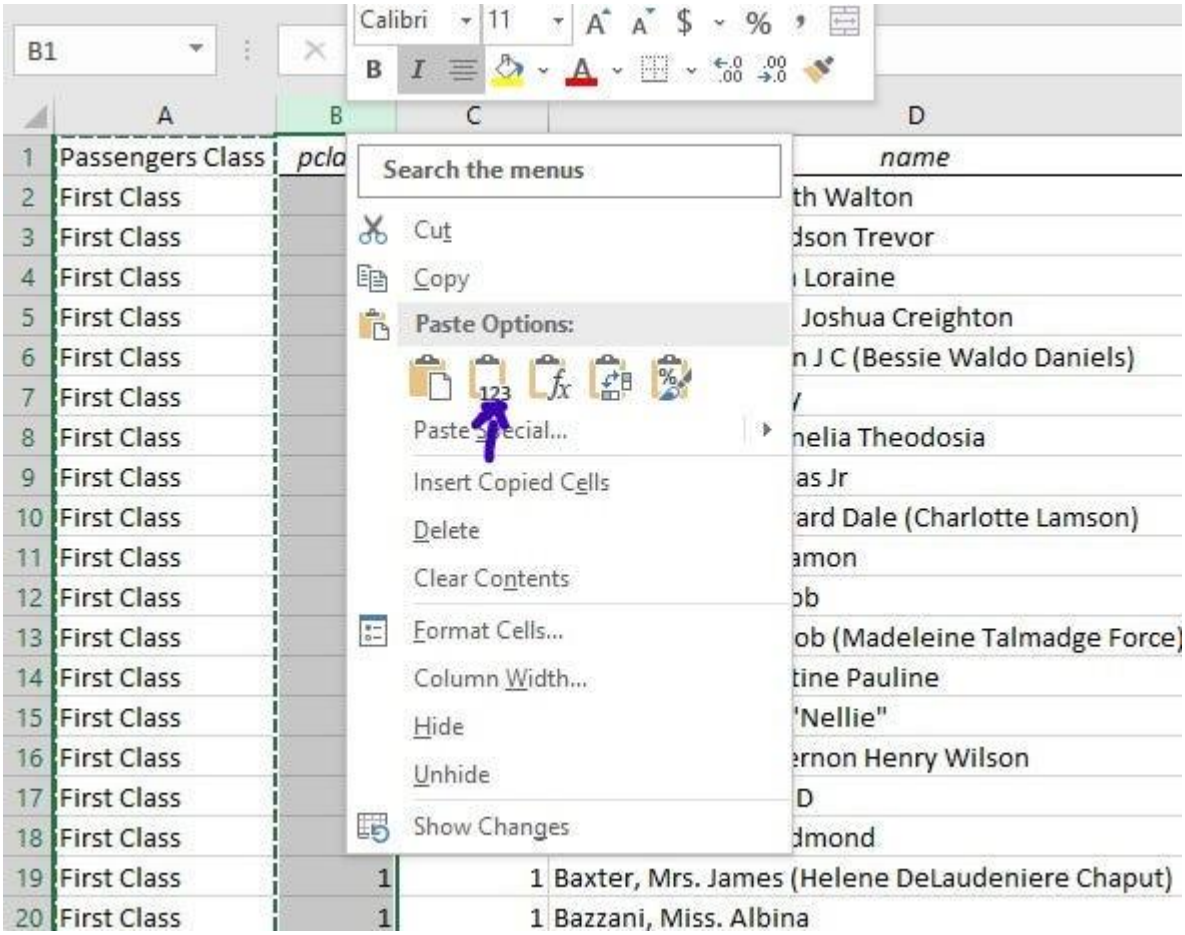
representations of First Class, Second Class, and Third Class. This was done by creating a new column called "Passenger Class", using the IF function to convert the numerical values to textual representations.

Once the new column was created, I copied the textual representations of passenger class and pasted them into the P Class column, using the "Paste Values" option available by right-clicking on the column. By using this method, we were able to represent the class of passengers in a more meaningful and easily understandable way.



	A	B	C	D	E	F	G	H	I
1	Passengers Class	*pclass*	*survived*	*name*	*sex*	*age*	*sibsp*	*parch*	*ticket*
2	First Class	1	1	Allen, Miss. Elisabeth Walton	female	29	0	0	24160
3	First Class	1	1	Allison, Master. Hudson Trevor	male	1	1	2	113781
4	First Class	1	0	Allison, Miss. Helen Loraine	female	2	1	2	113781
5	First Class	1	0	Allison, Mr. Hudson Joshua Creighton	male	30	1	2	113781

Once the new column was created, I copied the content of the "Passenger Class" column and pasted it into the "P Class" column, overwriting the original numerical values. To do this, I used the "Paste Values" option, which can be found by selecting the column you want to paste into and right-click. As shown below.



Survived Column

According to the data, the value '1' represents those who survived and the value '0' represents those who died. To make the data easier to understand, I used an IF Function to replace '1' with 'Yes' and '0' with 'No' in the 'Survived' column. As shown below.

```
=IF(C2:C1310=1,"Yes",IF(C2:C1310=0,"No"))
```

Then, I copied the new 'Yes' and 'No' column and pasted the values onto the 'Survived' column to replace the original '0' and '1' values, just like I did for the 'P Class' column. As shown below.

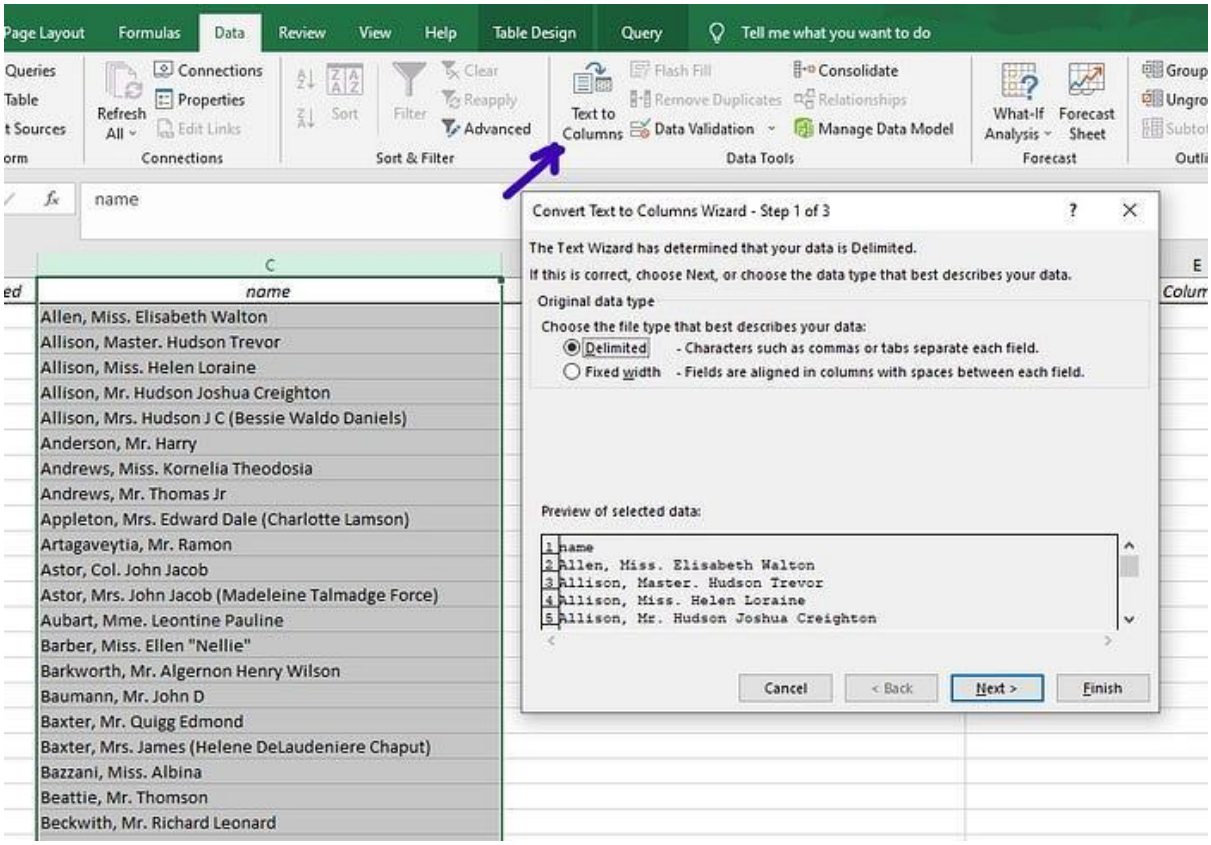
A	B
Passengers Class	*Survived*
First Class	Yes
First Class	Yes
First Class	No
First Class	No
First Class	No
First Class	Yes
First Class	Yes
First Class	No
First Class	Yes
First Class	No
First Class	No
First Class	Yes
First Class	Yes
First Class	Yes
First Class	Yes
First Class	No
First Class	No
First Class	Yes
First Class	Yes
First Class	No
First Class	Yes

Name Column

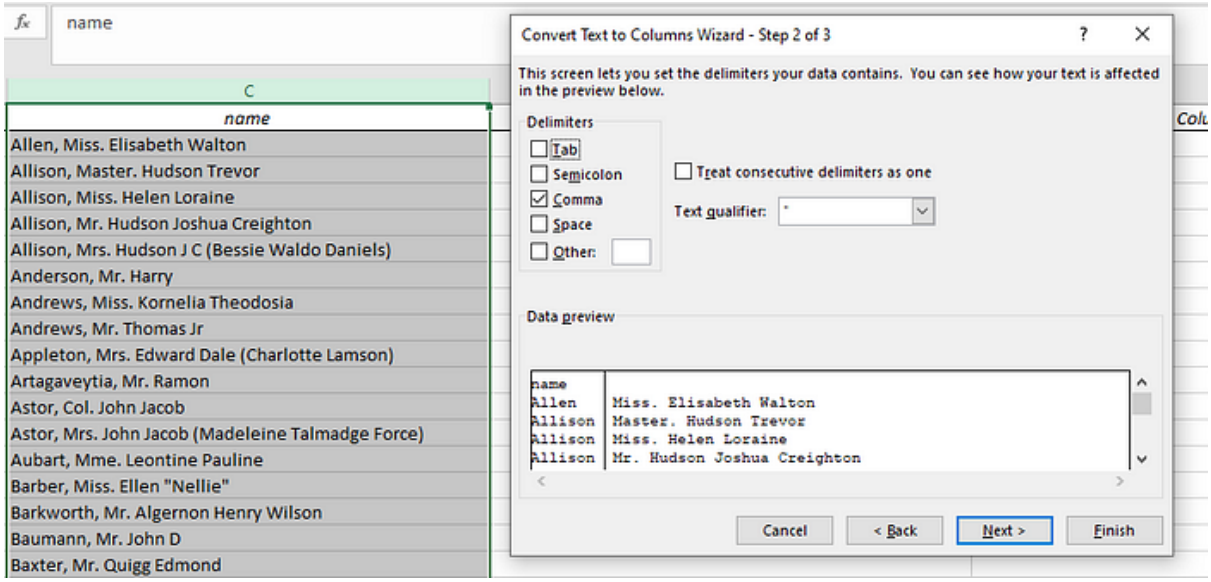
As shown below, upon uploading the raw data, it became apparent that the 'Name' column was disorganized, containing many inconsistencies. This prompted me to clean it up.

name
Allen, Miss. Elisabeth Walton
Allison, Master. Hudson Trevor
Allison, Miss. Helen Loraine
Allison, Mr. Hudson Joshua Creighton
Allison, Mrs. Hudson J C (Bessie Waldo Daniels)
Anderson, Mr. Harry
Andrews, Miss. Kornelia Theodosia
Andrews, Mr. Thomas Jr
Appleton, Mrs. Edward Dale (Charlotte Lamson)
Artagaveytia, Mr. Ramon
Astor, Col. John Jacob
Astor, Mrs. John Jacob (Madeleine Talmadge Force)
Aubart, Mme. Leontine Pauline
Barber, Miss. Ellen "Nellie"
Barkworth, Mr. Algernon Henry Wilson
Baumann, Mr. John D
Baxter, Mr. Quigg Edmond
Baxter, Mrs. James (Helene DeLaudeniere Chaput)
Bazzani, Miss. Albina
Beattie, Mr. Thomson
Beckwith, Mr. Richard Leonard

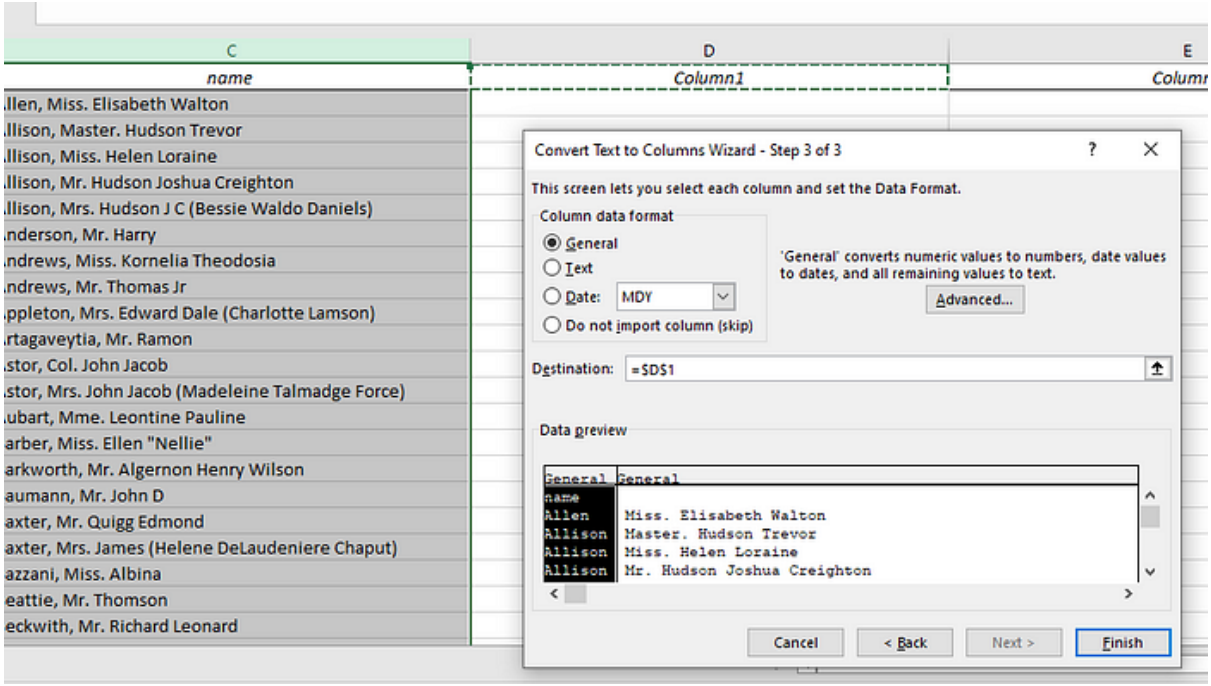
To start, I inserted two helper columns and then used the Delimiter option to separate the comma symbol from the names. I clicked on ***"Text to columns***" found in the ***Data*** ribbon. As shown below.



After that, I clicked delimiter. Then, “next” which took me to the next step as shown below.



I unselected the tab, chose Comma, and clicked “Next”. On the third wizard, I chose one of the new helper columns as the destination and clicked finish.

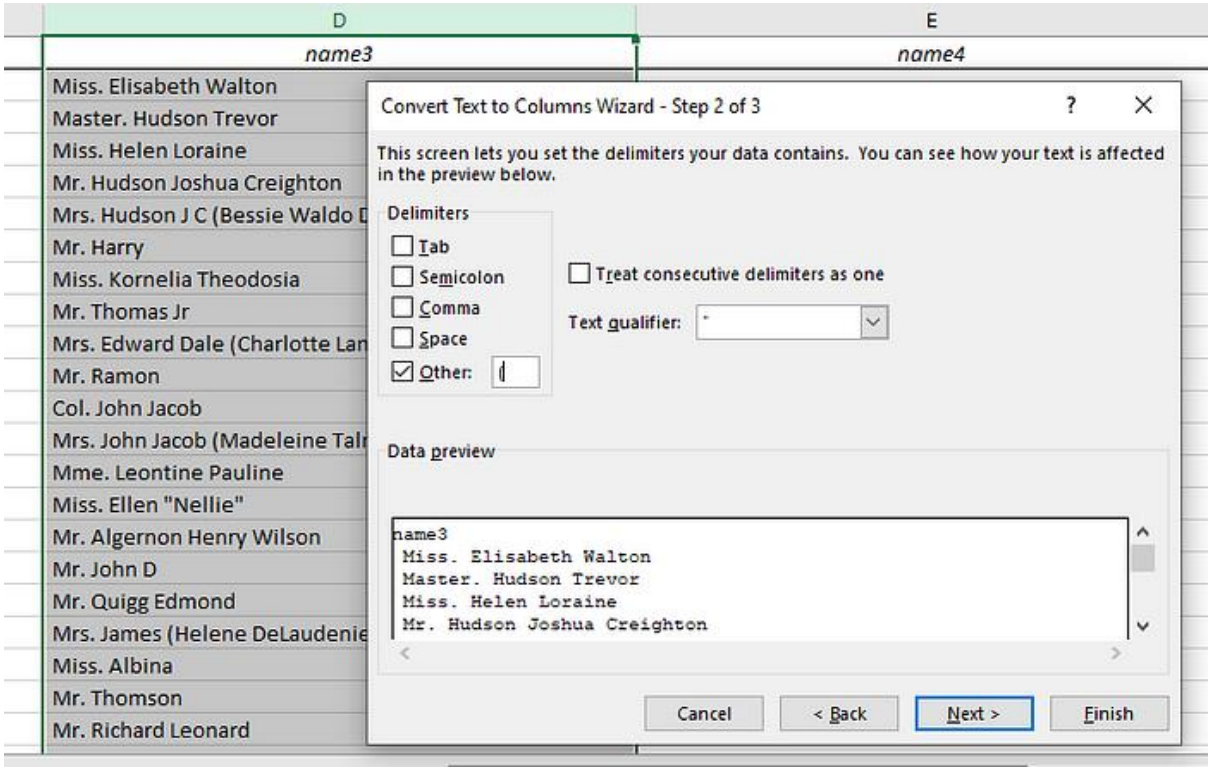


Below is the result of the wizard setup.

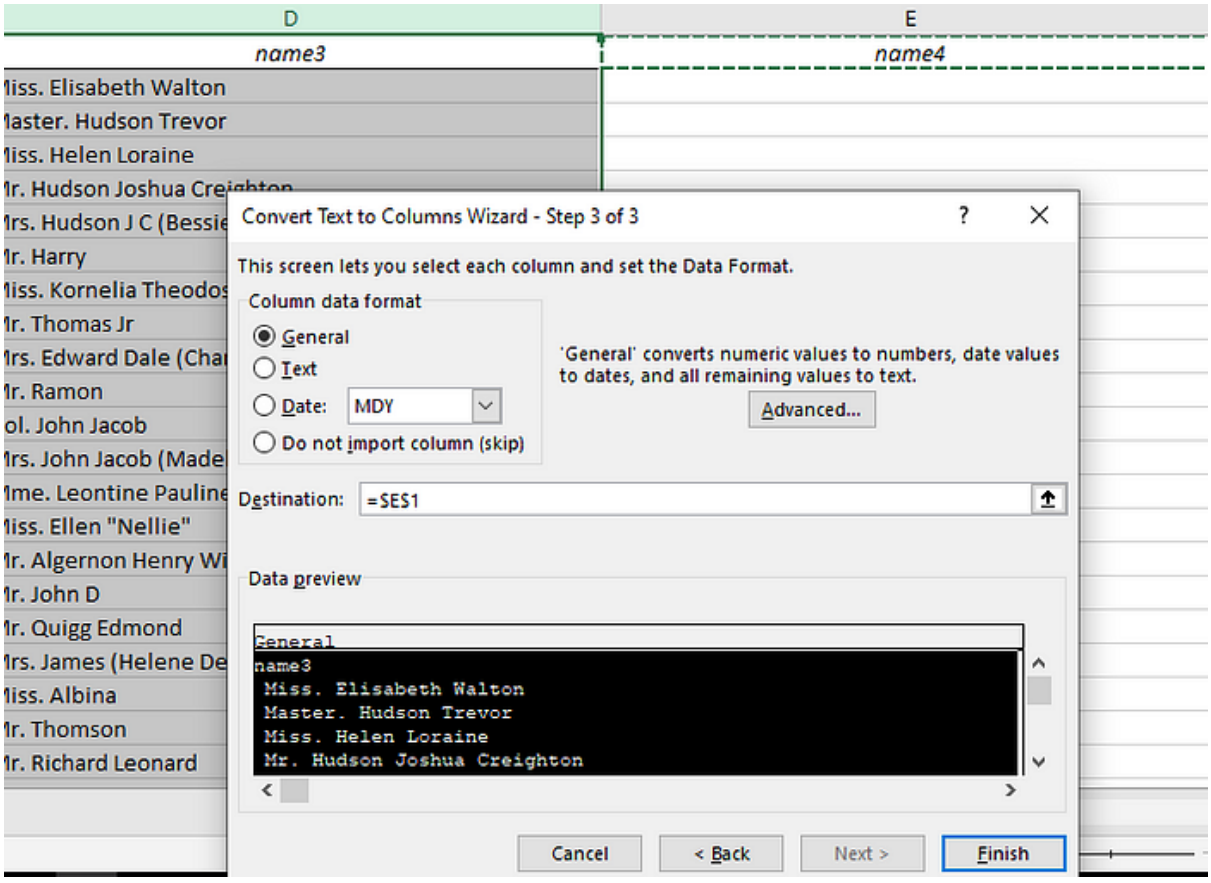
D	E
name2	*name3*
Allen	Miss. Elisabeth Walton
Allison	Master. Hudson Trevor
Allison	Miss. Helen Loraine
Allison	Mr. Hudson Joshua Creighton
Allison	Mrs. Hudson J C (Bessie Waldo Daniels)
Anderson	Mr. Harry
Andrews	Miss. Kornelia Theodosia
Andrews	Mr. Thomas Jr
Appleton	Mrs. Edward Dale (Charlotte Lamson)
Artagaveytia	Mr. Ramon
Astor	Col. John Jacob
Astor	Mrs. John Jacob (Madeleine Talmadge Force)
Aubart	Mme. Leontine Pauline
Barber	Miss. Ellen "Nellie"
Barkworth	Mr. Algernon Henry Wilson
Baumann	Mr. John D
Baxter	Mr. Quigg Edmond
Baxter	Mrs. James (Helene DeLaudeniere Chaput)
Bazzani	Miss. Albina
Beattie	Mr. Thomson
Beckwith	Mr. Richard Leonard

The above image shows that name 2 is clean, but name 3 still looks dirty. To clean column 3, I used the same ***Delimiter*** option and inserted another helper column. I highlighted column 3 and clicked on ***"Text to column"*** once more.

This time, I clicked on "Other" and inputed the open parenthesis symbol "(" as shown below.



I used the new helper column (name 4) as the destination. As shown below.



I also used the same procedure to clean the quotation marks (") and separated the names to work with only the first and last names.

	D	E	F	G	H	I	J
1	*name2*	*name322*	*name3222*	*name3223*	*name3224*	*name3225*	*name32*
2	Allen	Miss. Elisabeth Walton		Miss.	Elisabeth	Walton	
3	Allison	Master. Hudson Trevor		Master.	Hudson	Trevor	
4	Allison	Miss. Helen Loraine		Miss.	Helen	Loraine	
5	Allison	Mr. Hudson Joshua Creighton		Mr.	Hudson	Joshua	Creighto
6	Allison	Mrs. Hudson J C		Mrs.	Hudson	J	C
7	Anderson	Mr. Harry		Mr.	Harry		
8	Andrews	Miss. Kornelia Theodosia		Miss.	Kornelia	Theodosia	
9	Andrews	Mr. Thomas Jr		Mr.	Thomas	Jr	
10	Appleton	Mrs. Edward Dale		Mrs.	Edward	Dale	
11	Artagaveytia	Mr. Ramon		Mr.	Ramon		
12	Astor	Col. John Jacob		Col.	John	Jacob	
13	Astor	Mrs. John Jacob		Mrs.	John	Jacob	
14	Aubart	Mme. Leontine Pauline		Mme.	Leontine	Pauline	
15	Barber	Miss. Ellen		Miss.	Ellen		
16	Barkworth	Mr. Algernon Henry Wilson		Mr.	Algernon	Henry	Wilson
17	Baumann	Mr. John D		Mr.	John	D	
18	Baxter	Mr. Quigg Edmond		Mr.	Quigg	Edmond	
19	Baxter	Mrs. James		Mrs.	James		
20	Bazzani	Miss. Albina		Miss.	Albina		
21	Beattie	Mr. Thomson		Mr.	Thomson		
22	Beckwith	Mr. Richard Leonard		Mr.	Richard	Leonard	

I used "Text to column" (space) to separate the names

I then joined them together using the **Concatenate function. As shown below.**

```
=CONCATENATE(C2, " ", D2, " ", B2)
```

	B	C	D	E
	name2			
	Allen	Miss.	Elisabeth	Miss. Elisabeth Allen
	Allison	Master.	Hudson	Master. Hudson Allison
	Allison	Miss.	Helen	Miss. Helen Allison

Finally, I copied the combined names in that column and clicked ***"paste values***" on the original name column. I renamed the name column as **"Full name".** As shown below.

C
Full Name
Miss. Elisabeth Allen
Master. Hudson Allison
Miss. Helen Allison
Mr. Hudson Allison
Mrs. Hudson Allison
Mr. Harry Anderson
Miss. Kornelia Andrews
Mr. Thomas Andrews
Mrs. Edward Appleton
Mr. Ramon Artagaveytia
Col. John Astor
Mrs. John Astor
Mme. Leontine Aubart
Miss. Ellen Barber
Mr. Algernon Barkworth
Mr. John Baumann
Mr. Quigg Baxter
Mrs. James Baxter
Miss. Albina Bazzani
Mr. Thomson Beattie
Mr. Richard Beckwith

The cleaned name column

Headers

Later on, I realized that the headers were inconsistent. For instance, they were written in improper cases. Hence, I used the ***Proper function*** to clean the names. Then, I dragged it to the last row so it could fill. I also copied and pasted it on the main header row.

```
=PROPER(A2)
```

Sex Column

The 'Proper Function' was used to capitalize the first letter of each word in the 'sex' column which was not too dirty.

Column1	Sex
Female	female
Male	male
Female	female
Male	male
Female	female
Male	male
Female	female
Male	male
Female	female
Male	male
Male	male
Female	female
Female	female
Female	female
Male	male
Male	male
Male	male
Female	female
Female	female
Male	male
Male	male

Family Size

In order to better understand the data, I decided to create a new column called "Family Size" which was calculated by adding the values of the "Sibsp" (Siblings/Spouses) and "Parch" (Parent/Children) columns.

To achieve this, I used the ***"SUM" function*** and entered "=SUM(F2, G2)" in the formula bar.

F	G	H
Siblings/Spouses	Parents/Children	Family Size
0	0	0
1	2	3
1	2	3
1	2	3
1	2	3
0	0	0
1	0	1
0	0	0
2	0	2
0	0	0
1	0	1
1	0	1
0	0	0

Family Description

In order to better understand the family size of each passenger, I added a new column to the data set that indicates whether or not a passenger had family members on board. I used the ***If Function*** to classify the family size of each passenger, labeling those that had no family members on board as "No family" and those with at least one family member as "Family".

Family Size	Family Description
0	No Family
0	No Family
0	No Family
0	No Family
5	Family
5	Family
5	Family
5	Family
5	Family
5	Family
0	No Family
0	No Family
2	Family
1	Family
1	Family
2	Family
2	Family
2	Family
0	No Family
1	Family
1	Family
0	No Family

Additionally, I formatted the Ticket column as text in order to make it easier to read and analyze.

Ticket
24160
113781
113781
113781
113781
19952
13502
112050
11769
PC 17609
PC 17757
PC 17757
PC 17477
19877
27042
PC 17318
PC 17558

Fare

Based on my assessment, I found that the records were mostly accurate. However, a few passengers were listed as having paid zero fare. I assumed that these individuals were part of the cabin crew, including the captain, officers, and deckhands. As the crew members were responsible for navigating the vessel, maintaining its equipment and systems, and ensuring its safety and seaworthiness, it made sense that they would not need to pay. To make the fare column clearer, I formatted it to display two decimal places and added the dollar sign as shown below.

K
Fare ($)
$211.34
$151.55
$151.55
$151.55
$151.55
$26.55
$77.96
$0.00
$51.48
$49.50
$227.53
$227.53
$69.30
$78.85
$30.00
$25.93

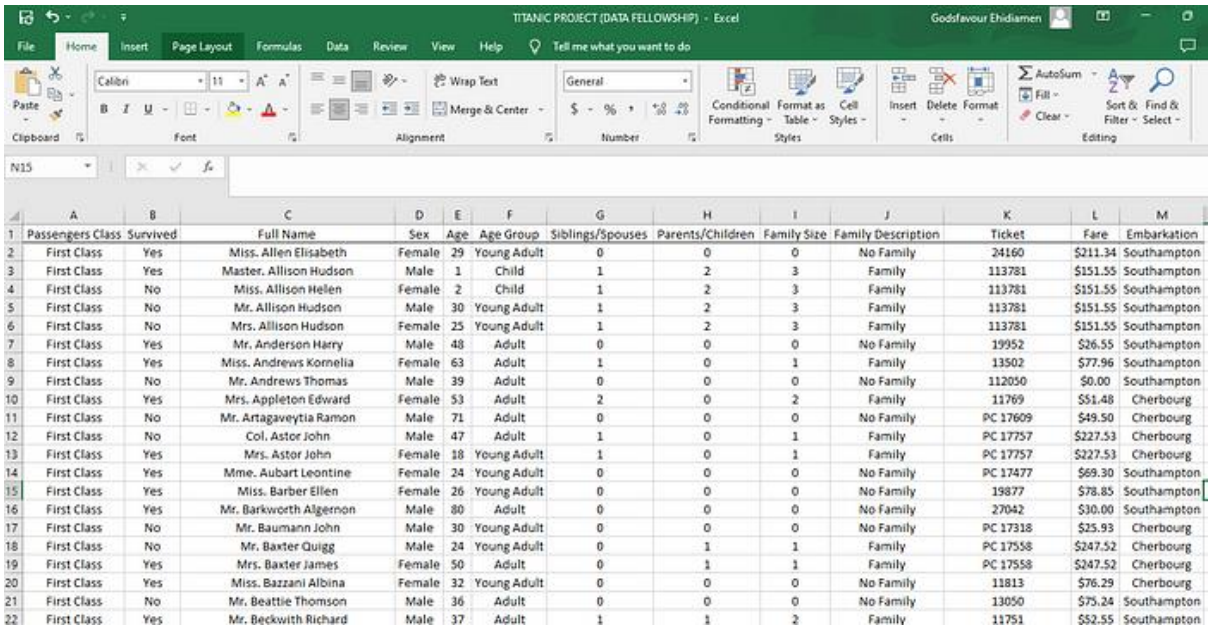
The fare column

Embarked

The Titanic dataset included a column called "Embarked" which provided information about the port of embarkation for each passenger. The dataset includes three ports of embarkation, namely Cherbourg (C), Queenstown (Q), and Southampton (S). To make the data more understandable, I replaced each letter with the corresponding names of the ports.

```
=IF(L2="S","Southampton",IF(L2="C","Cherbourg",IF(L2="Q","Queenstown")))
```

IF function to replace letters with their real words



Passengers Class	Survived	Full Name	Sex	Age	Age Group	Siblings/Spouses	Parents/Children	Family Size	Family Description	Ticket	Fare	Embarkation
First Class	Yes	Miss. Allen Elisabeth	Female	29	Young Adult	0	0	0	No Family	24160	$211.34	Southampton
First Class	Yes	Master. Allison Hudson	Male	1	Child	1	2	3	Family	113781	$151.55	Southampton
First Class	No	Miss. Allison Helen	Female	2	Child	1	2	3	Family	113781	$151.55	Southampton
First Class	No	Mr. Allison Hudson	Male	30	Young Adult	1	2	3	Family	113781	$151.55	Southampton
First Class	No	Mrs. Allison Hudson	Female	25	Young Adult	1	2	3	Family	113781	$151.55	Southampton
First Class	Yes	Mr. Anderson Harry	Male	48	Adult	0	0	0	No Family	19952	$26.55	Southampton
First Class	Yes	Miss. Andrews Kornelia	Female	63	Adult	1	0	1	Family	13502	$77.96	Southampton
First Class	No	Mr. Andrews Thomas	Male	39	Adult	0	0	0	No Family	112050	$0.00	Southampton
First Class	Yes	Mrs. Appleton Edward	Female	53	Adult	2	0	2	Family	11769	$51.48	Cherbourg
First Class	No	Mr. Artagaveytia Ramon	Male	71	Adult	0	0	0	No Family	PC 17609	$49.50	Cherbourg
First Class	No	Col. Astor John	Male	47	Adult	1	0	1	Family	PC 17757	$227.53	Cherbourg
First Class	Yes	Mrs. Astor John	Female	18	Young Adult	1	0	1	Family	PC 17757	$227.53	Cherbourg
First Class	Yes	Mme. Aubart Leontine	Female	24	Young Adult	0	0	0	No Family	PC 17477	$69.30	Southampton
First Class	Yes	Miss. Barber Ellen	Female	26	Young Adult	0	0	0	No Family	19877	$78.85	Southampton
First Class	Yes	Mr. Barkworth Algernon	Male	80	Adult	0	0	0	No Family	27042	$30.00	Southampton
First Class	No	Mr. Baumann John	Male	30	Young Adult	0	0	0	No Family	PC 17318	$25.93	Cherbourg
First Class	No	Mr. Baxter Quigg	Male	24	Young Adult	0	1	1	Family	PC 17558	$247.52	Cherbourg
First Class	Yes	Mrs. Baxter James	Female	50	Adult	0	1	1	Family	PC 17558	$247.52	Cherbourg
First Class	Yes	Miss. Bazzani Albina	Female	32	Young Adult	0	0	0	No Family	11813	$76.29	Cherbourg
First Class	No	Mr. Beattie Thomson	Male	36	Adult	0	0	0	No Family	13050	$75.24	Southampton
First Class	Yes	Mr. Beckwith Richard	Male	37	Adult	1	1	2	Family	11751	$52.55	Southampton

After cleaning

Now that you understand how I cleaned the data, continue reading so you can gain a deeper insight into the analysis. 🤓

STEP 2: DATA ANALYSIS

After perfectly cleaning the data, I conducted a thorough analysis.

I determined the survival rate by gender, passenger class, and family size.

Survival Rate of Passengers:

From the analysis, only 38.30% of passengers survived while 61.80% died.

Average Age of Passengers:

The average age of passengers is 30

Proportion Of Male And Female Passengers:

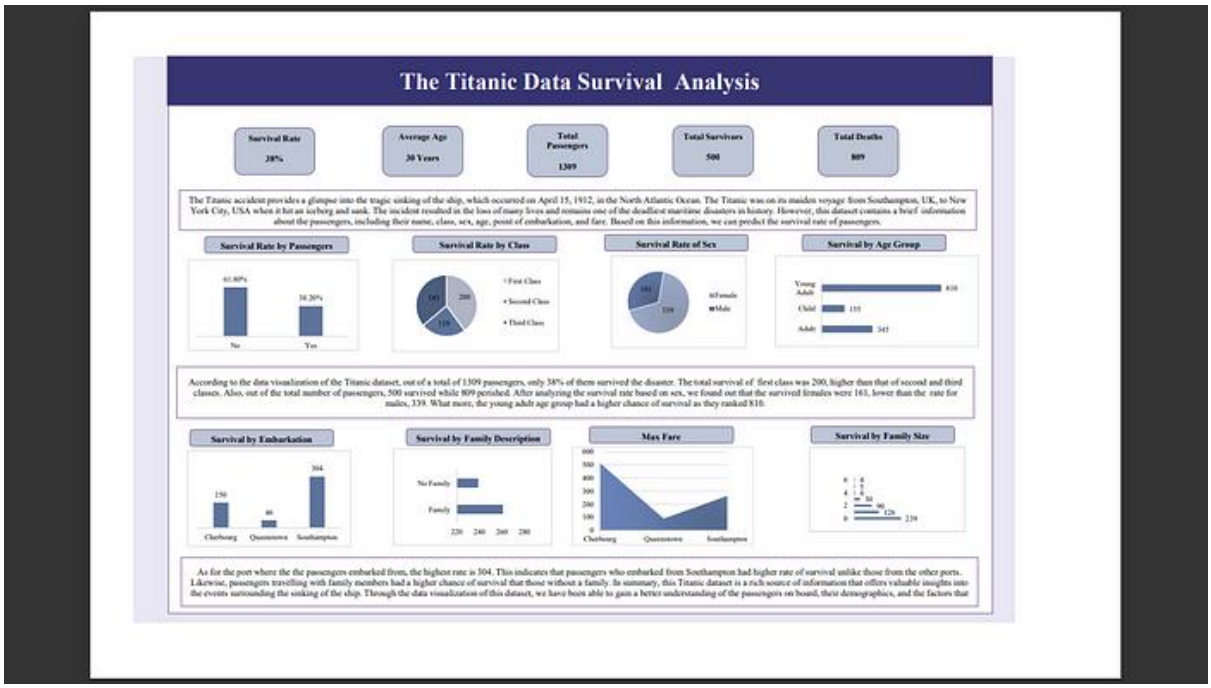
I noticed that there were more females (64.40%) than males (35.60%) on the ship.

Passengers With The Highest Survival Rate:

The first class passengers had the highest survival rate of 40%. The second and third classes had 23.80% and 36.20%, respectively.

	A	B	C	D	E	F	G	H	I	J
2										
3	**Row Labels**	**Count of Survived**		**Row Labels**	**Count of Survived**		**Average of Age**			**Row Labels**
4	No	61.80%		No	809		30			Adult
5	Yes	38.20%		Yes	500					Child
6	**Grand Total**	**100.00%**		**Grand Total**	**1309**					Young Adult
7										**Grand Total**
8										
9										
10	Proportion of male and female passengers			Survived	Yes		Survived	Yes		**Row Labe**
11										Cherbourg
12	**Row Labels**	**Count of Sex**		**Row Labels**	**Count of Passengers Class**		**Row Labels**	**Count of Sex**		Queenstown
13	Female	35.60%		First Class	40.00%		Female	339		Southampton
14	Male	64.40%		Second Class	23.80%		Male	161		**Grand Total**
15	**Grand Total**	**100.00%**		Third Class	36.20%		**Grand Total**	**500**		
16				**Grand Total**	**100.00%**					
17										
18										
19										
20	Survived	Yes		Survived	No		Survived	Yes		Survived
21										
22	**Row Labels**	**Count of Passengers Class**		**Row Labels**	**Count of Sex**		**Row Labels**	**Count of Embarked**		**Row Labe**
23	First Class	200		Female	127		Cherbourg	150		0

STEP 3: DATA VISUALIZATION



Data visualization

CONCLUSION

In summary, the Titanic dataset provides a wealth of information that sheds light on the circumstances surrounding the ship's tragic sinking. By visualizing this dataset, I was able to gain a better understanding of the passengers on board, including their demographics, as well as the factors that played a role in determining their survival during the disaster..

Thanks for reading.