@YT 2023

Microsoft Excel Mini Task

Created By : Yulindra Tita W.

LIST OF CONTENT

01 What is Microsoft Excel?

Microsoft Excel is an application that allows you to enter data in writing, text, numbers, or other complex functions. This application is also known as a spreadsheet. It helps users to calculate, analyze, and present data. The main function of the Microsoft Excel program is as a number processing application. In addition, there are other functions of Microsoft Excel, wich are:

- **Create, edit, sort, analyse, and summarize data.**
- **Do arithmetic and statistical calculations.**
- **Helps solve logical and mathematical problems.**
- **Create various charts and diagrams (data vizualisation).**
- **Create financial records, budgets, and financial reports.**
- **Calculate and manage investments, loans, sales, inventories, etc.**

02 Data Description

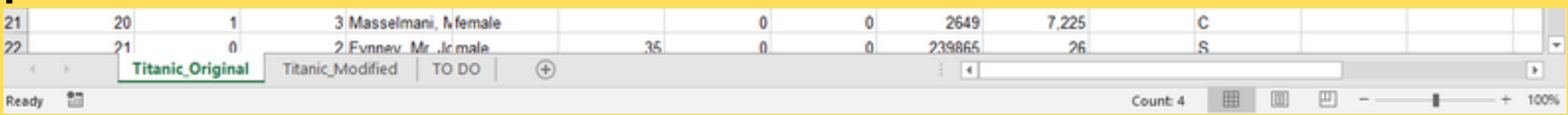
This popular open-source dataset offers information on the passengers onboard the Titanic ship when it sank on April 15, 1912. It can be used by data analytics beginners interested in data cleaning and preprocessing, descriptive statistics, data visualization and predictive modeling. Some of the variables included in the dataset:

- PassengerId - A unique identifier for each passenger.
- Survived - This shows whether the passenger survived or not
- Pclass - A passenger's class (1 = 1st, 2 = 2nd, 3 = 3rd).
- Name - A passenger's name.
- Sex - A passenger's gender.
- Age - A passenger's age.
- SibSp - The number of siblings/spouses aboard.
- Parch - The number of parents/children aboard.
- Ticket - The ticket number.
- Fare - The fare paid for the ticket.
- Cabin - The cabin number.
- Embarked - The port of embarkation (C = Cherbourg, Q = Queenstown, S = Southampton).

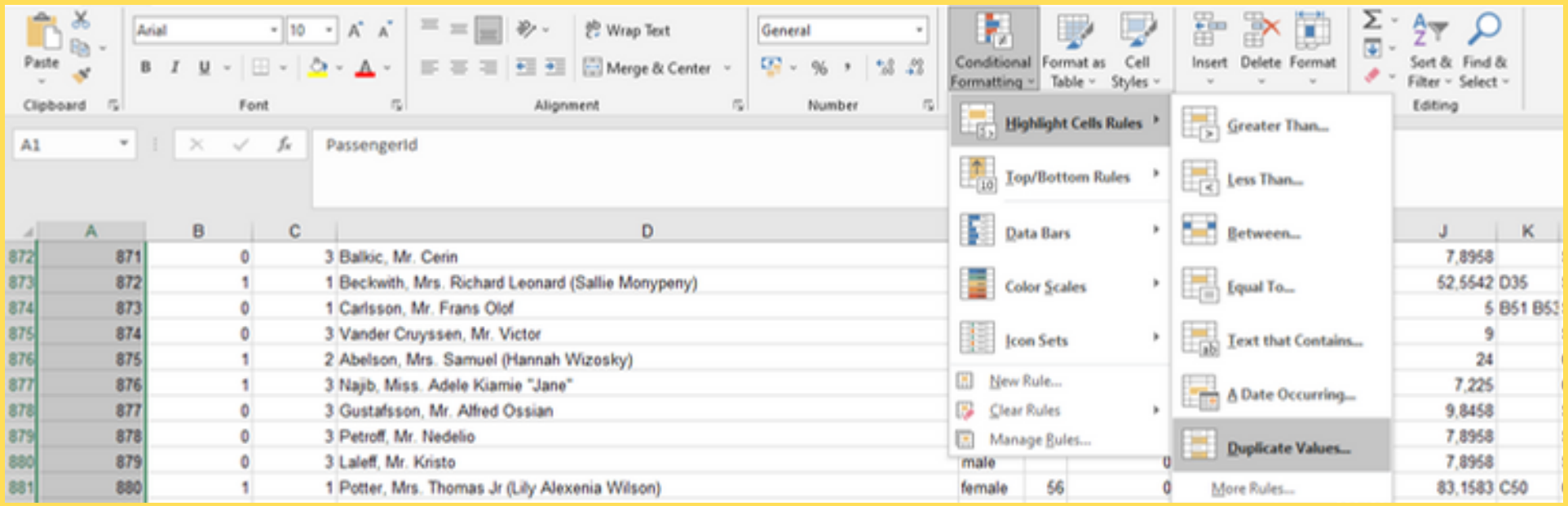
Data source:
https://hackernoon.com/15-excel-datasets-for-data-analytics-beginners

03 Data Cleaning

Before processing the data, I duplicate the titanic dataset into a new sheet so as not to change the primary dataset and make it easier to process the data.



- **First, check for duplicate data. If there is duplicate data, we have to delete one of them. Use the conditional formatting, highlight cells rules, then duplicate values. Before using conditional formatting, we have to block the column that contains unique data. In this section, the column "PassengerId" refers to unique data.**



- **Based on the data description, we know that column "*survived*" is only contains "*0*" and "*1*". The data cannot be understood by the level managers who will use the data, so the data must be converted to "*yes*" and "*no*" statements. To convert the data, we use *IF* statement on the formula: =IF(Titanic_Original!$B2=0;"No";"Yes"). It's mean if "0" then fill "*No*" else "*Yes*".**

Before

Survived
0
1
1
1
0
0
0

After

Survived
No
Yes
Yes
Yes
No
No
No

- **The next step is to add the ordinal number description to the pclass column, where 1 becomes 1st, 2 becomes 2nd, 3 becomes 3rd with the formula: (IFS(Titanic_Original!$C2=1;CONCAT(Titanic_Original!$C2;"st");Titanic_Original!$C2=2;CONCAT(Titanic_Original!$C2;"nd");Titanic_Original!$C2=3;CONCAT(Titanic_Original!$C2;"rd")))**

Before

Pclass
3
1
3
1
3
3
3

After

Pclass
3rd
1st
3rd
1st
3rd
3rd
1st

@YT 2023

- **Fill the blank column with the number 0 on the cabin column and for the column age we mantain the blank column with it's median. Use the formula "IF IsBlank" an "Median" as follows,**

1. **Column age:**
   =IF(ISBLANK(Titanic_Original!$F2);MEDIAN(Titanic_Original!$F$2:$F$892);IF(Titanic_Original!$F2<1;1;Titanic_Original!$F2))
2. **Column cabin:**
   =IF(ISBLANK(Titanic_Original!K2);0;Titanic_Original!K2)

Before

Age	Cabin
22	
38	C85
26	
35	C123
35	

After

Age	Cabin
22	0
38	C85
26	0
35	C123
35	0
28	0

- **Change the header from "Sex" to "Gender" dan formats gender column, where the gender description begins with a capital letter using the formula: =PROPER(Titanic_Original!E2).**

Before

Gender
male
female
female
female

After

Gender
Male
Female
Female
Female
Male

@YT 2023

- **Formatted the embarkation column to display more complete information, if "C" then “Cherboug”, if "S" then “Southampton”, If "Q" then “Queenstown”, and when the data is blank fill with "unknown". The formula should be: =IFS(Titanic_Original!L2="C";"Cherbourg"; Titanic_Original!L2="Q";"Queenstown";Titanic_Original!L2="S";"Southampton";ISBLANK(Titanic_Original!L2);"Unknown").**

Before
Embarked
S
C
S
S
S
Q

After
Embarked
Southampton
Cherbourg
Southampton
Southampton
Southampton
Queenstown

- **The next step after the data formatting process is completed is to make the data into a table by blocking the entire data and then clicking sortcut keyboard Ctrl+K.**

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket
1	No	3rd	Braund, Mr. Owen Harris	Male	22	1	0	A/5 211
2	Yes	1st	Cumings, Mrs. John Bradley (Floren	Female	38	1	0	PC 175
3	Yes	3rd	Heikkinen, Miss. Laina	Female	26	0	0	STON/(
4	Yes	1st	Futrelle, Mrs. Jacques Heath (Lily M	Female	35	1	0	1138
5	No	3rd	Allen, Mr. William Henry	Male	35	0	0	3734
6	No	3rd	Moran, Mr. James	Male	0	0	0	3308
7	No	1st	McCarthy, Mr. Timothy J	Male	54	0	0	174
8	No	3rd	Palsson, Master. Gosta Leonard	iginal!E9)	2	3	1	3499
9	Yes	3rd	Johnson, Mrs. Oscar W (Elisabeth \	Female	27	0	2	3477
10	Yes	2nd	Nasser, Mrs. Nicholas (Adele Achen	Female	14	1	0	2377

@YT 2023

04 Pivot Table & Data Visualization

Pivot Table is an Excel feature that is commonly used to manage and display data in a practical way, so users can convert data sets into an easy-to-read table. Before the data can be visualized, the pivot table must be defined first. In this case study, we're going to show some data taken from the dataset, namely: total passengers: alive and dead, total passengers by gender, passenger trends by age, survivors by embarked, and passengers by class.
Data visualization is the process of using visual elements such as diagrams, graphs, or maps to represent data. To make data visualization more dynamic, you need to add a slicer to the dashboard. Slicers act as data filters on Pivot Tables. Here are some charts used in this case study:

- Bar chart: To see the difference in quantity, proportion, or frequency between categories clearly,
- Donut chart: used to show the percentage A circle graph allows us to clearly see the contribution of each part to the whole. A
- Line chart or line graph: is a plot chart that visualizes data changes. In a line chart, a data point is connected to a line, creating a pattern that depicts a trend or fluctuation.

@YT 2023

Count of PassengerId	Gender		
Embarked	Female	Male	Grand Total
Cherbourg	73	95	168
Queenstown	36	41	77
Southampton	203	441	644
Grand Total	**312**	**577**	**889**

Pivot table passenger by embarked

Survived	Count of PassengerId
Yes	38,38%
No	61,62%
Grand Total	**100,00%**

Pivot table survived percentage

Count of PassengerId	Gender		
Survived	Female	Male	Grand Total
Yes	233	109	342
No	81	468	549
Grand Total	**314**	**577**	**891**

Pivot table score card main information

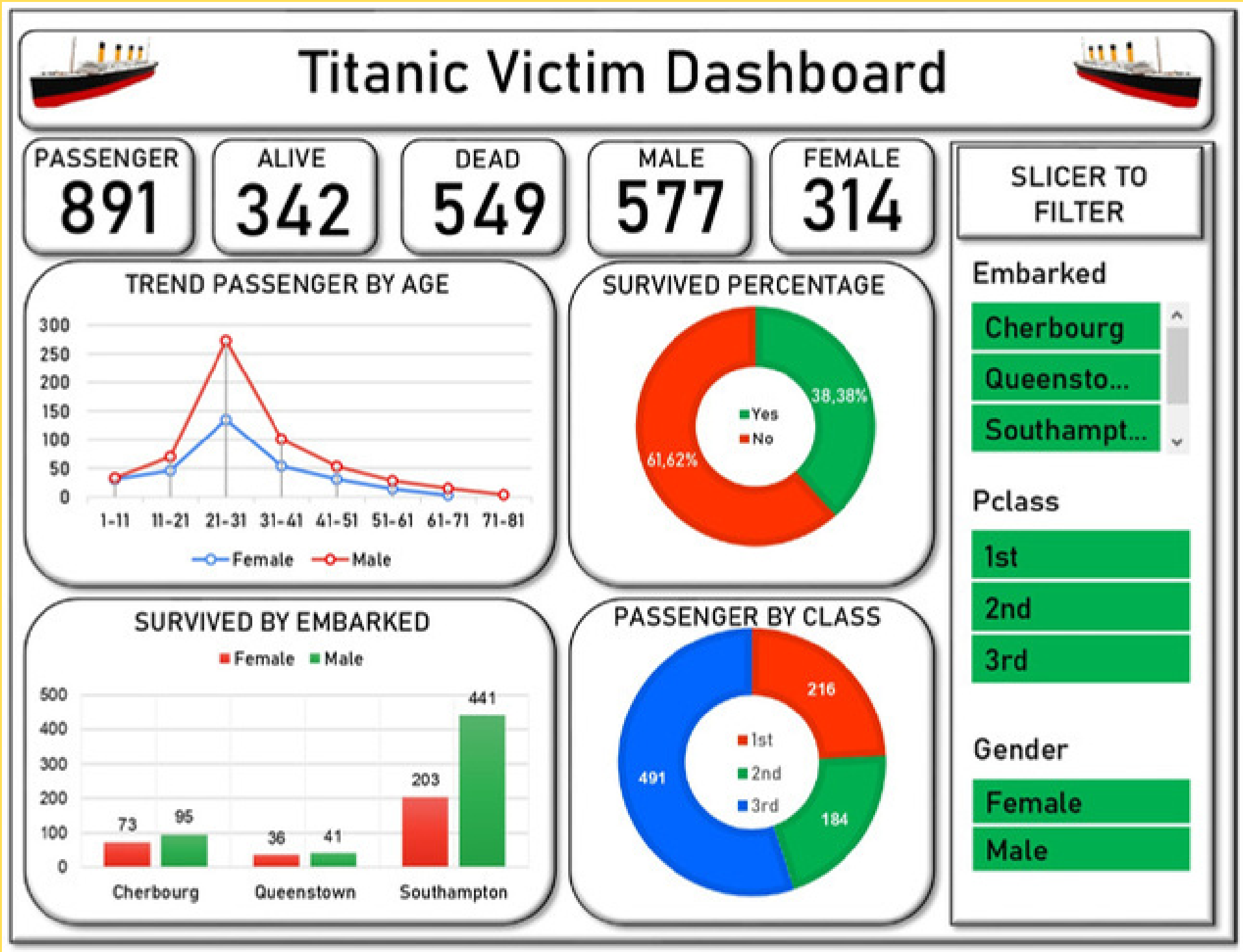
Count of PassengerId	Gender		
Age	Female	Male	Grand Total
1-11	31	33	64
11-21	46	70	116
21-31	135	273	408
31-41	54	101	155
41-51	31	53	84
51-61	14	28	42
61-71	3	15	18
71-81		4	4
Grand Total	**314**	**577**	**891**

Pivot table trend passenger by age

Pclass	Count of Ticket
1st	216
2nd	184
3rd	491
Grand Total	**891**

Pivot table passenger by class

05 Dashboard



Let’s connect !

Yulindra Tita

yulindratita@gmail.com

Please like and share it!

Thanks for the support!