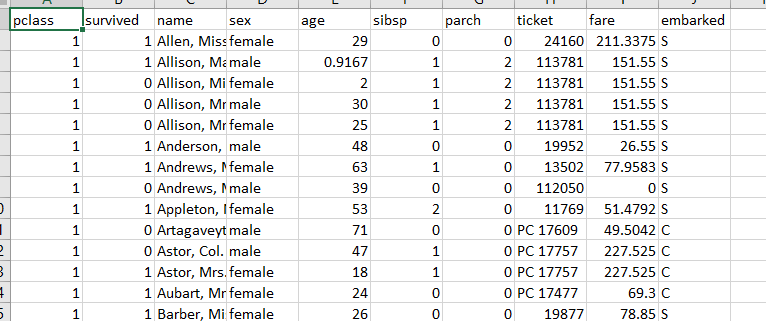
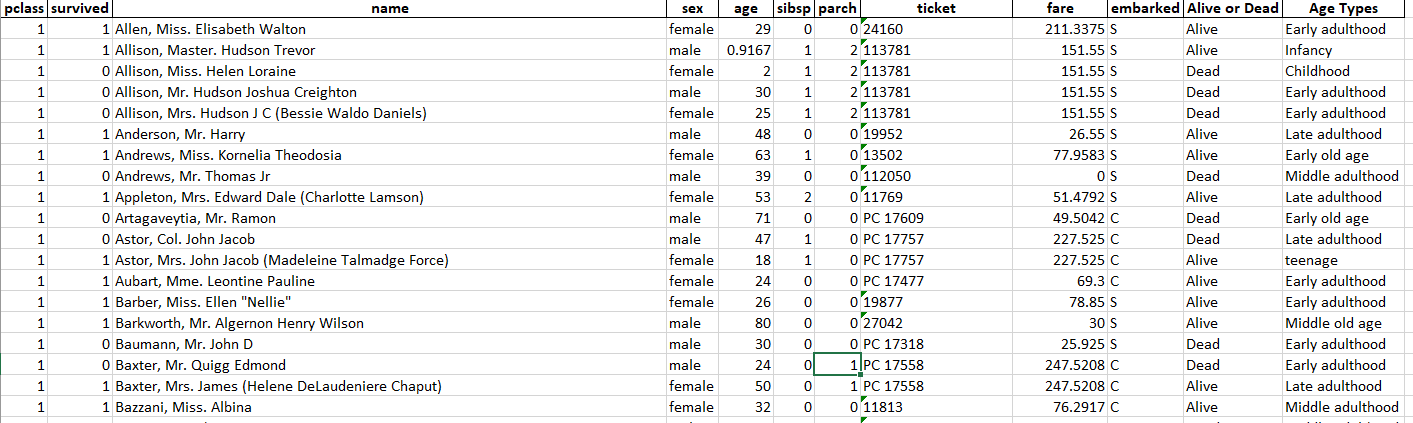
**Before Cleaning**





**After Cleaning**





**Cleaning the CSV & Excel File**

**By**

**using**

**Python**



import pandas as pd  
  
df = pd.read\_csv('titanic new.csv')  
  
print(df.isnull().sum())



import pandas as pd  
  
df = pd.read\_csv('titanic new.csv')  
temp = df  
temp['age'] = temp['age'].fillna(round(temp['age'].mean()))  
  
print(print(temp.isnull().sum()))



*"""  
1. Infancy (neonate and up to one year age)  
2. Childhood (1 to 11 years old)  
3. Adolescence or teenage (from 12 to 18 years old)  
4. Early adulthood (18 to 30 years)  
5. Middle adulthood (30 to 45 years)  
6. Late adulthood (45 to 60 years)  
7. Early old age (60 to 75 years)  
8. Middle old age (75 to 90 years)  
9. Late old age (over 90 years)  
"""*import pandas as pd  
import numpy as np  
  
df = pd.read\_excel('Final\_Data.xlsx')  
temp = df  
  
conditions = [  
 (temp['age'] <= 1),  
 (temp['age'] > 1) & (temp['age'] <= 11),  
 (temp['age'] > 11) & (temp['age'] <= 18),  
 (temp['age'] > 18) & (temp['age'] <= 30),  
 (temp['age'] > 30) & (temp['age'] <= 45),  
 (temp['age'] > 45) & (temp['age'] <= 60),  
 (temp['age'] > 60) & (temp['age'] <= 75),  
 (temp['age'] > 75) & (temp['age'] <= 90),  
 (df['age'] > 90)  
 ]  
  
values = ["Infancy",  
"Childhood","teenage",  
"Early adulthood",  
"Middle adulthood",  
"Late adulthood",  
"Early old age",  
"Middle old age",  
"Late old age"]  
  
temp['Age Types'] = np.select(conditions, values)  
  
print(temp)



import pandas as pd  
  
df = pd.read\_csv('titanic new.csv')  
temp = df  
temp['fare'] = temp['fare'].fillna((temp['fare'].mean()))  
  
print(print(temp.isnull().sum()))



import pandas as pd  
  
df = pd.read\_csv('titanic new.csv')  
temp = df  
temp['embarked'] = temp['embarked'].fillna("s")  
  
print(print(temp.isnull().sum()))



import pandas as pd  
  
df = pd.read\_csv('titanic new.csv')  
temp = df  
temp['Alive or Dead'] = ["Alive " if x ==1 else "Dead" for x in temp['survived']]  
  
print(temp)

Cleaning the give data by using Python



import pandas as pd  
import numpy as np  
  
df = pd.read\_csv('titanic new.csv')  
temp = df  
  
temp['fare'] = temp['fare'].fillna((temp['fare'].mean()))  
  
temp['embarked'] = temp['embarked'].fillna("s")  
  
temp['age'] = temp['age'].fillna(round(temp['age'].mean()))  
  
temp['Alive or Dead'] = ["Alive " if x ==1 else "Dead" for x in temp['survived']]  
  
conditions = [  
 (temp['age'] <= 1),  
 (temp['age'] > 1) & (temp['age'] <= 11),  
 (temp['age'] > 11) & (temp['age'] <= 18),  
 (temp['age'] > 18) & (temp['age'] <= 30),  
 (temp['age'] > 30) & (temp['age'] <= 45),  
 (temp['age'] > 45) & (temp['age'] <= 60),  
 (temp['age'] > 60) & (temp['age'] <= 75),  
 (temp['age'] > 75) & (temp['age'] <= 90),  
 (df['age'] > 90)  
 ]  
  
values = ["Infancy","Childhood","teenage","Early adulthood","Middle adulthood","Late adulthood","Early old age","Middle old age","Late old age"]  
  
temp['Age Types'] = np.select(conditions, values)  
  
df.to\_excel("Final\_Data.xlsx",index=False)



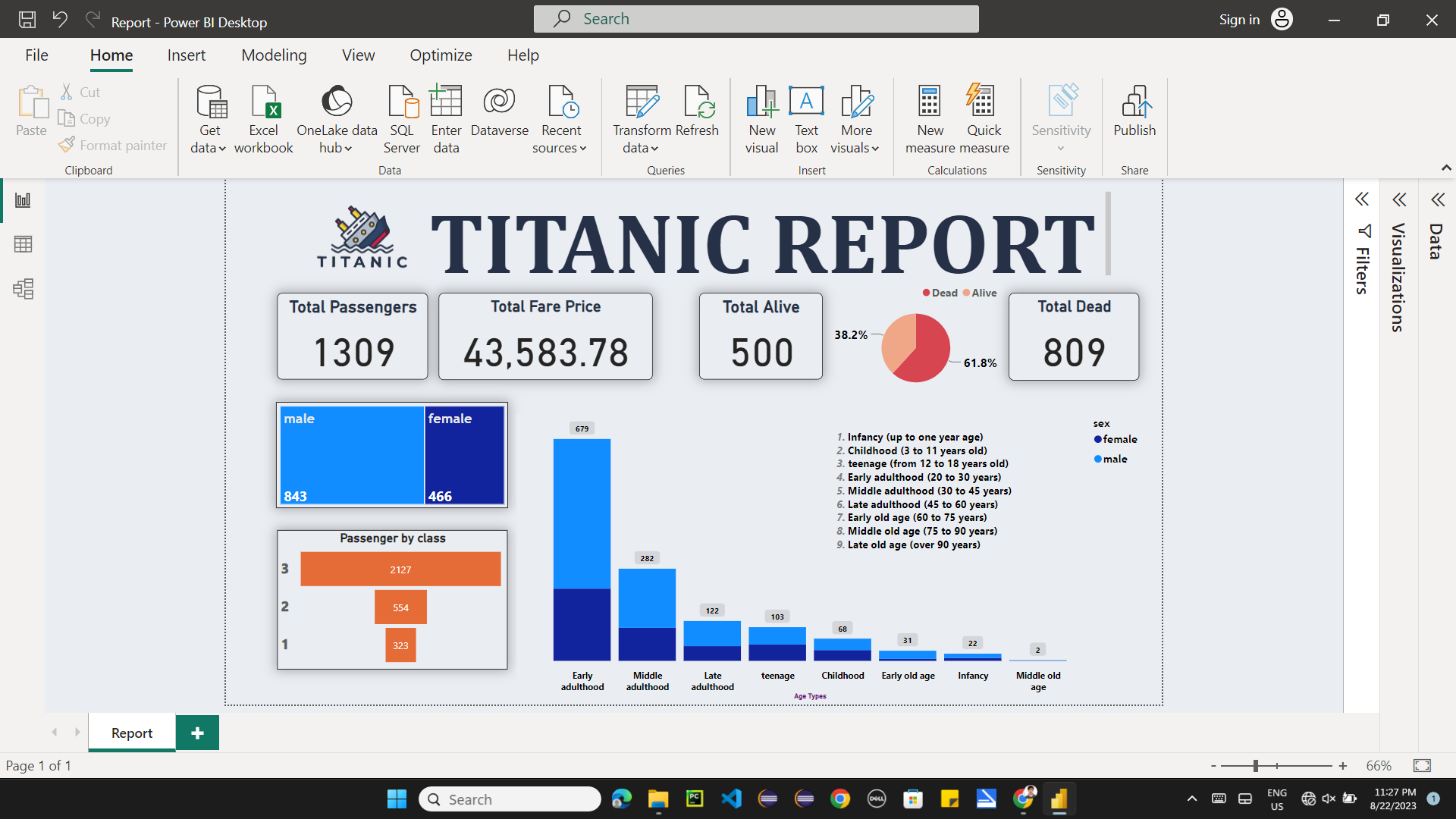




**Full Dashboard**

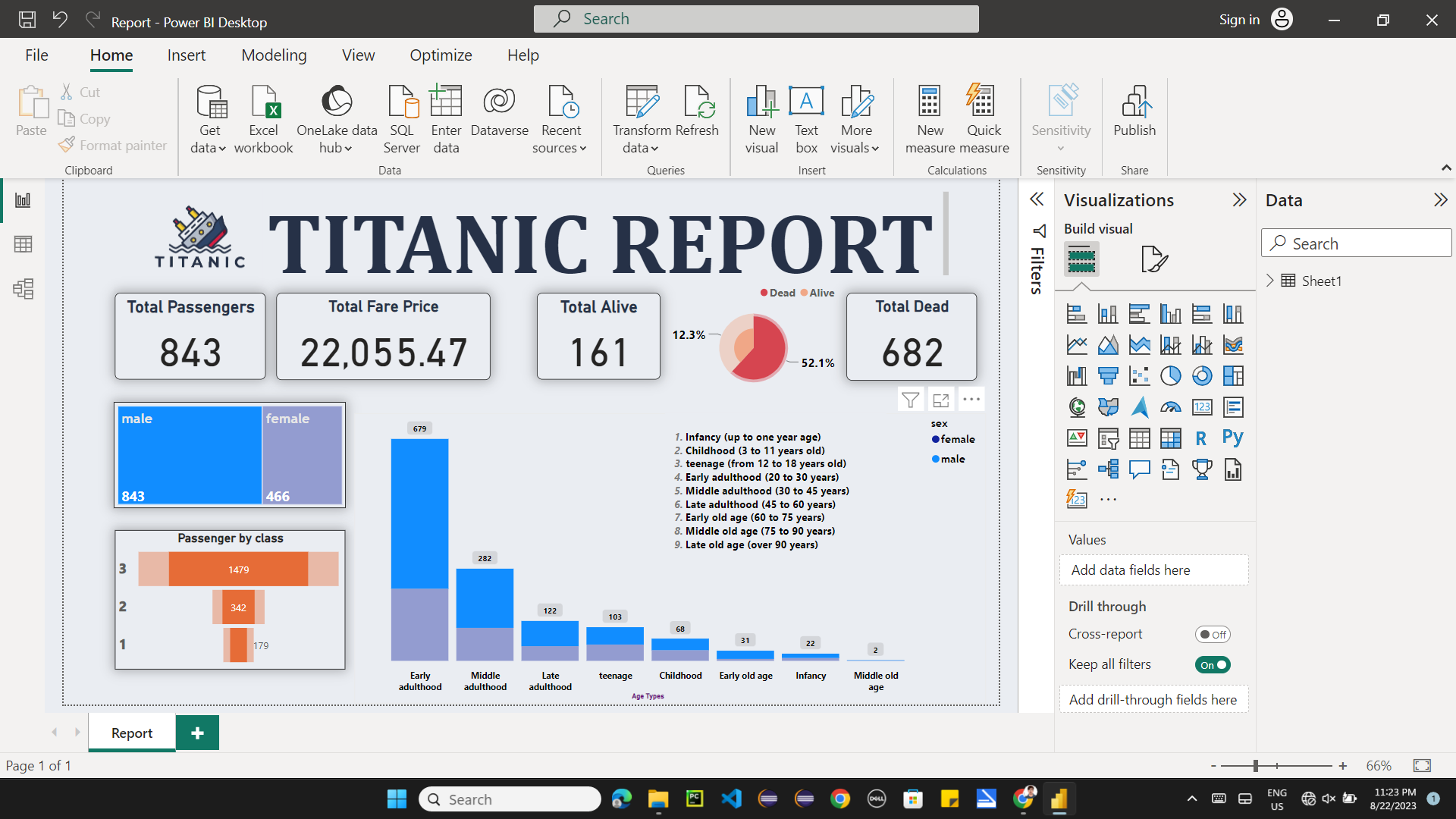
**by using**

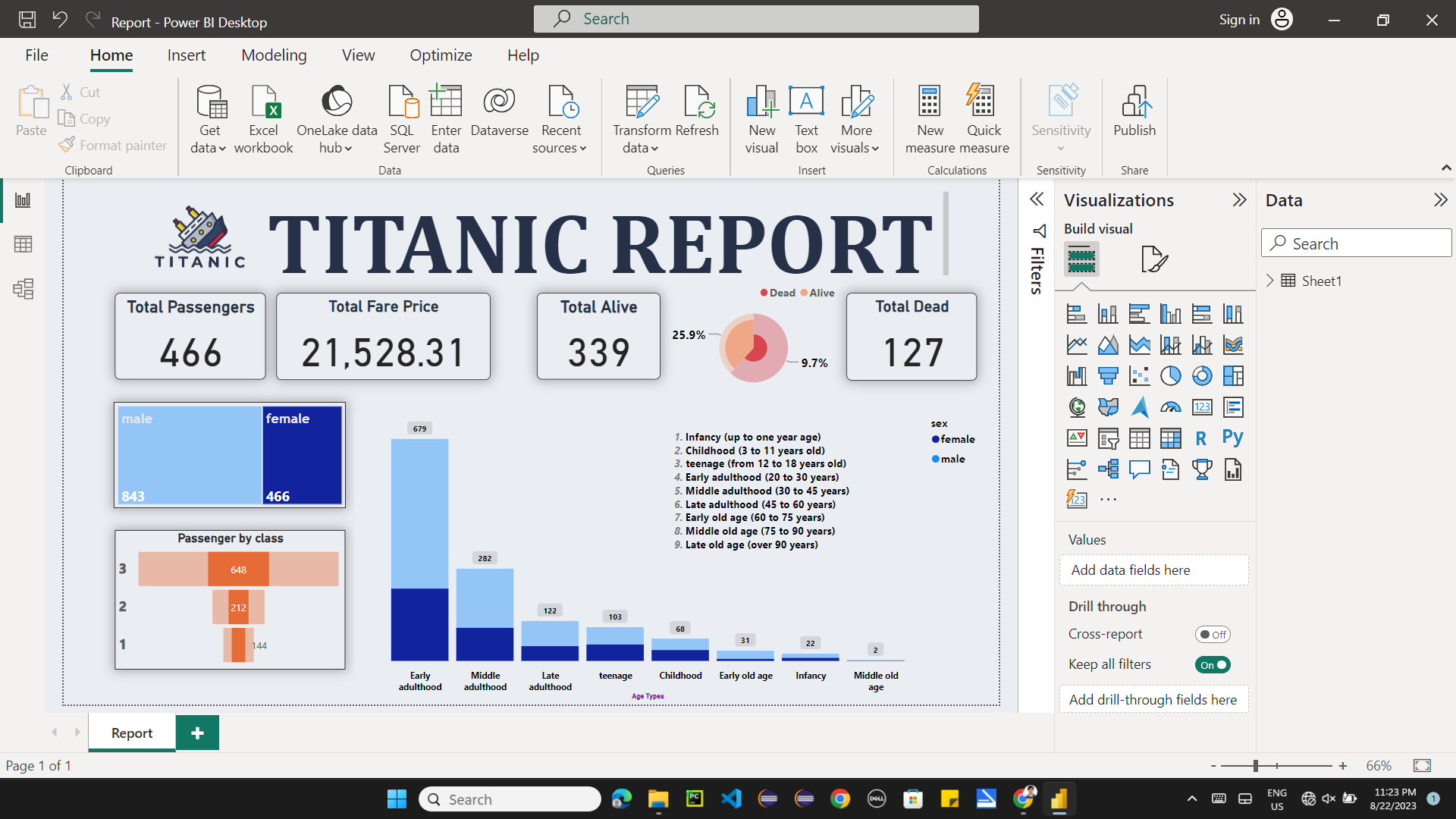
**Power BI**



**Classification based on**

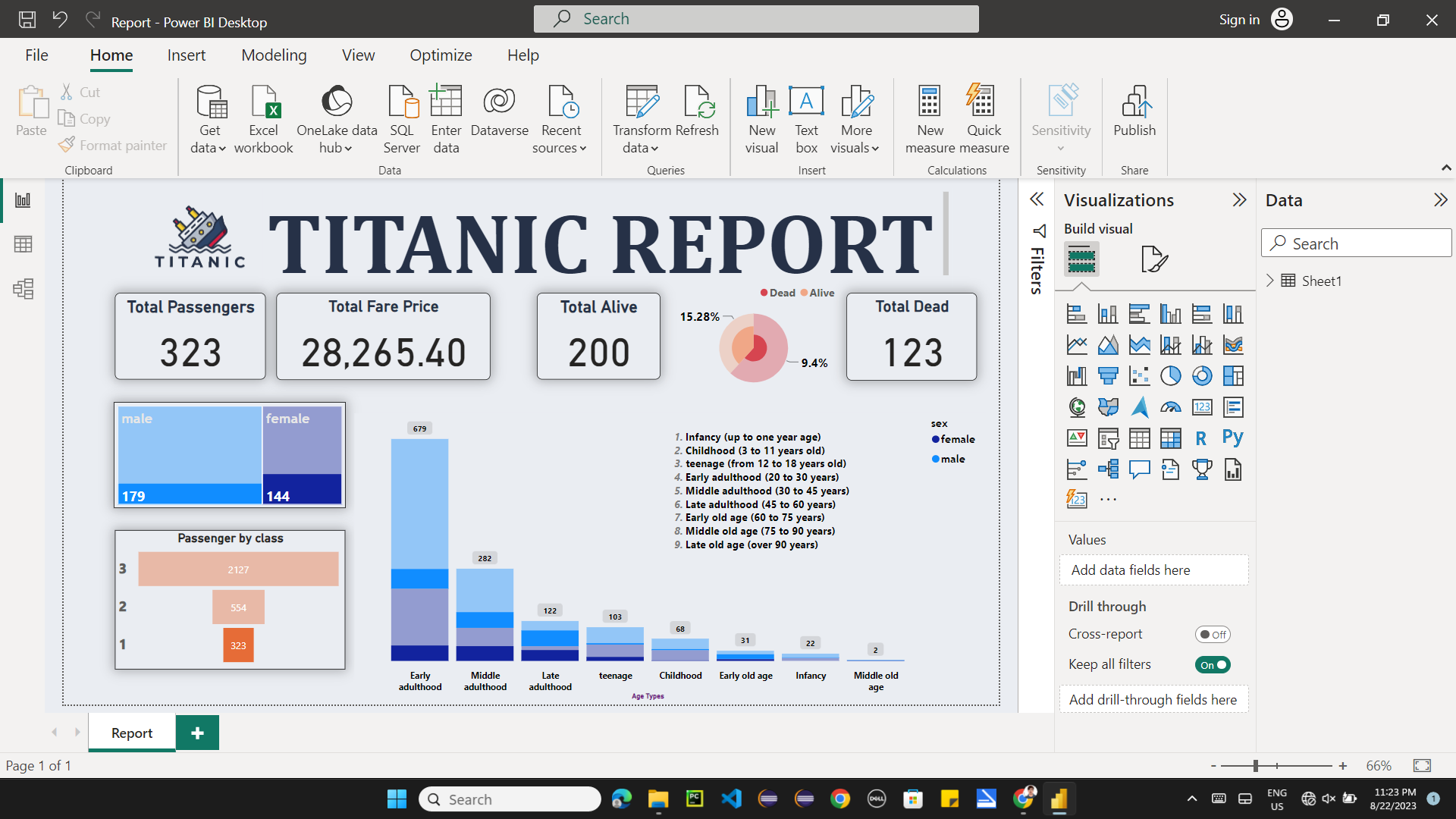
**Gender**

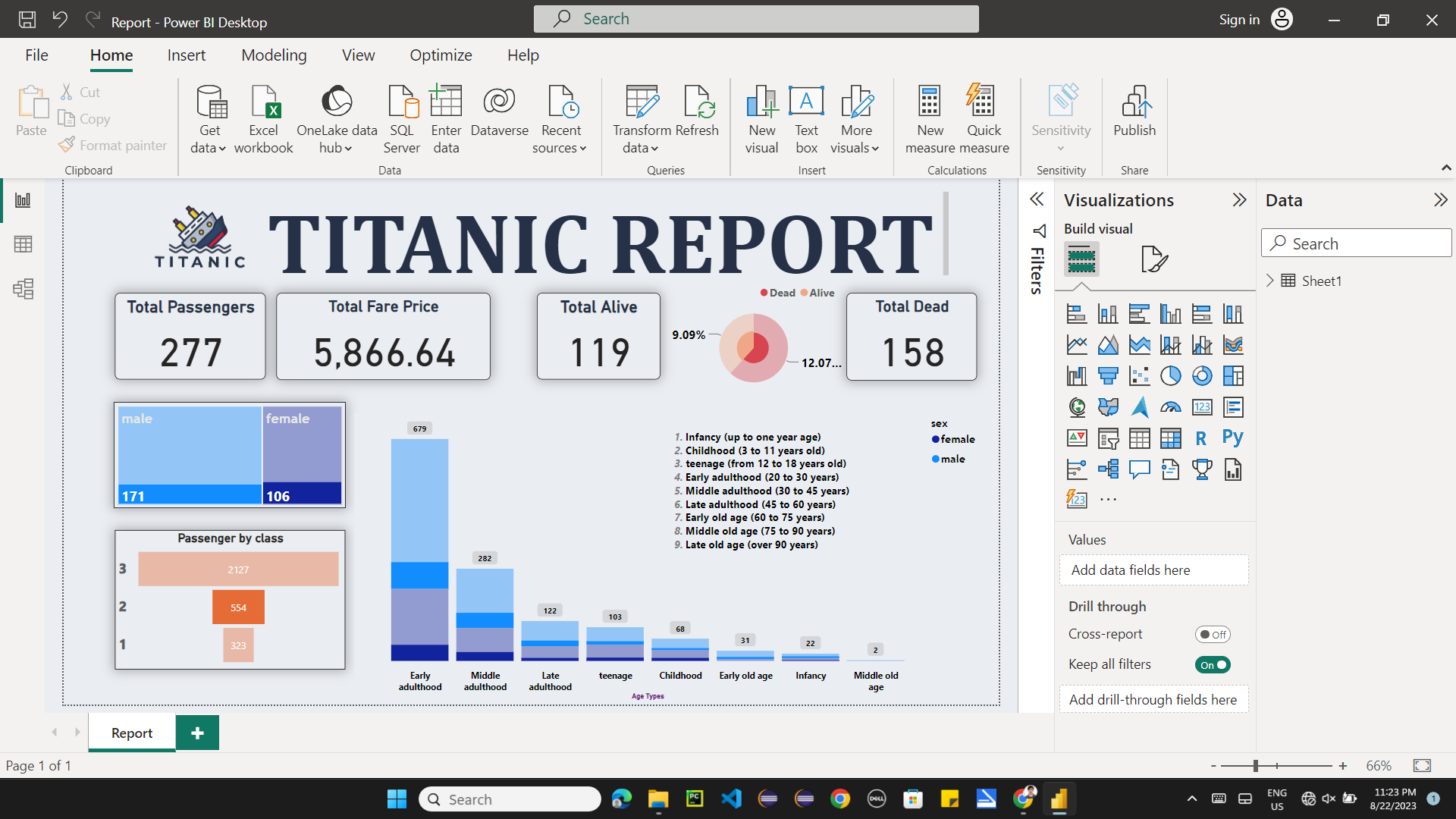


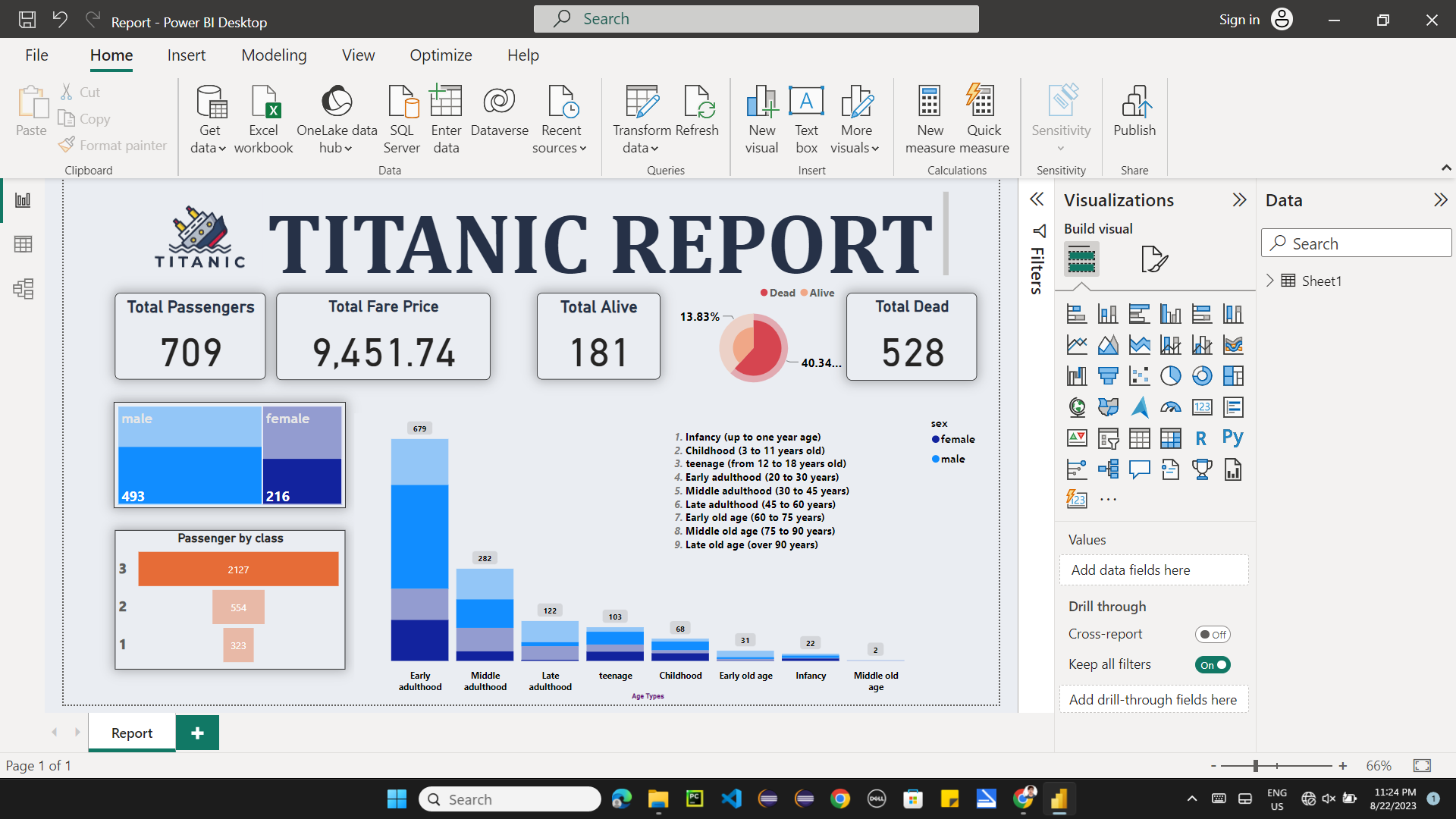


**Classification based on Classes**

**[1st, 2nd, 3rd]**



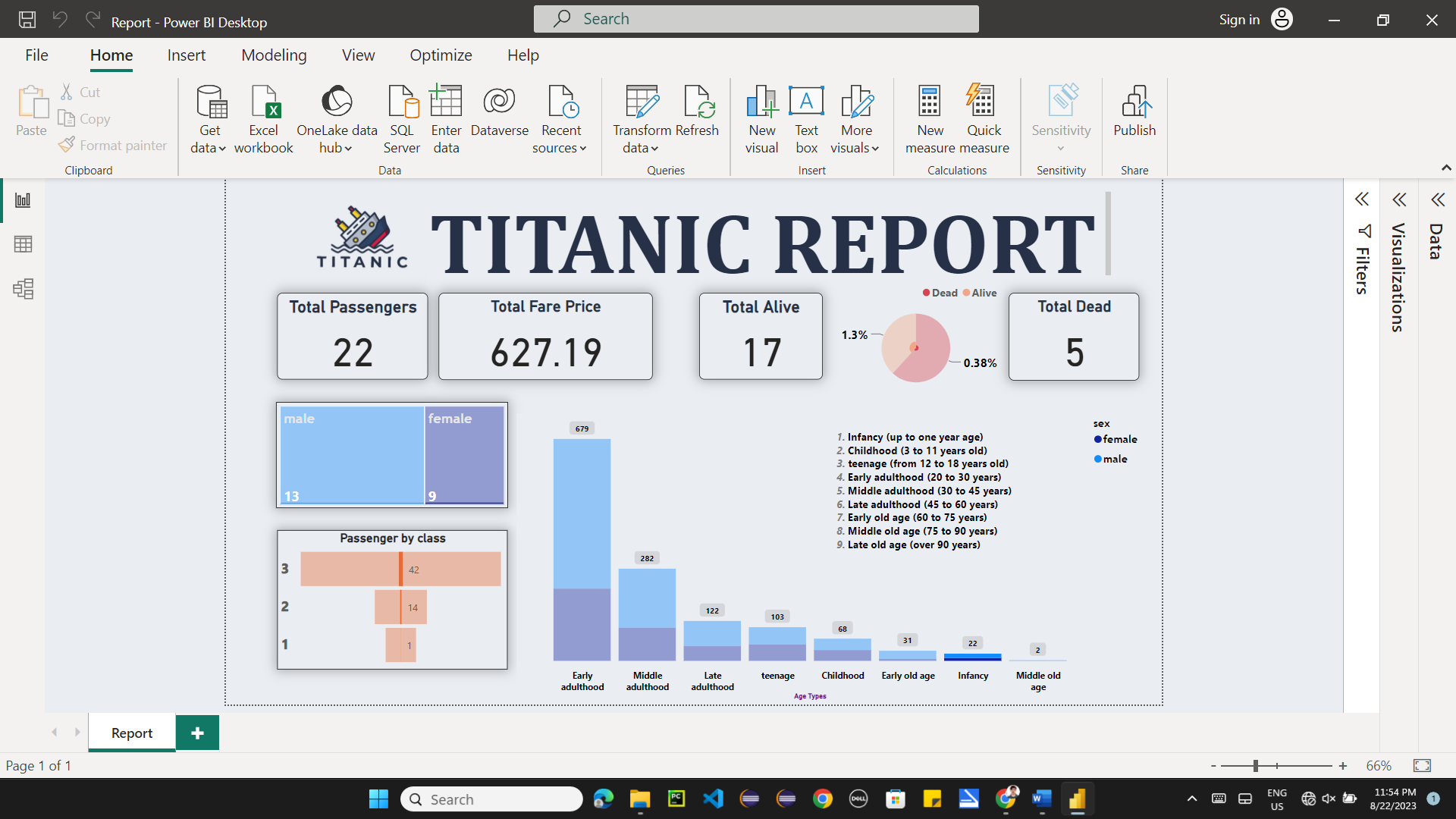


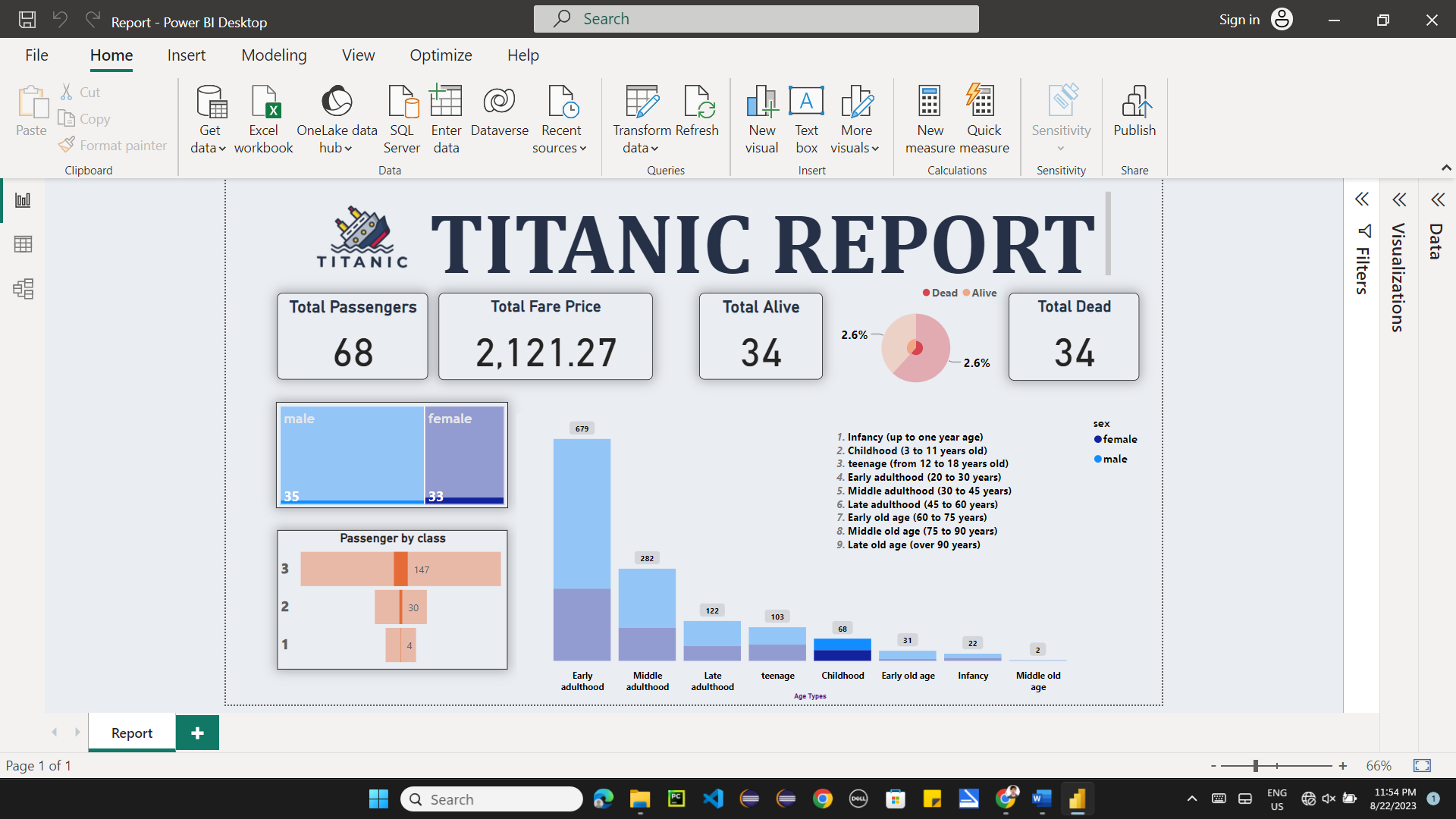


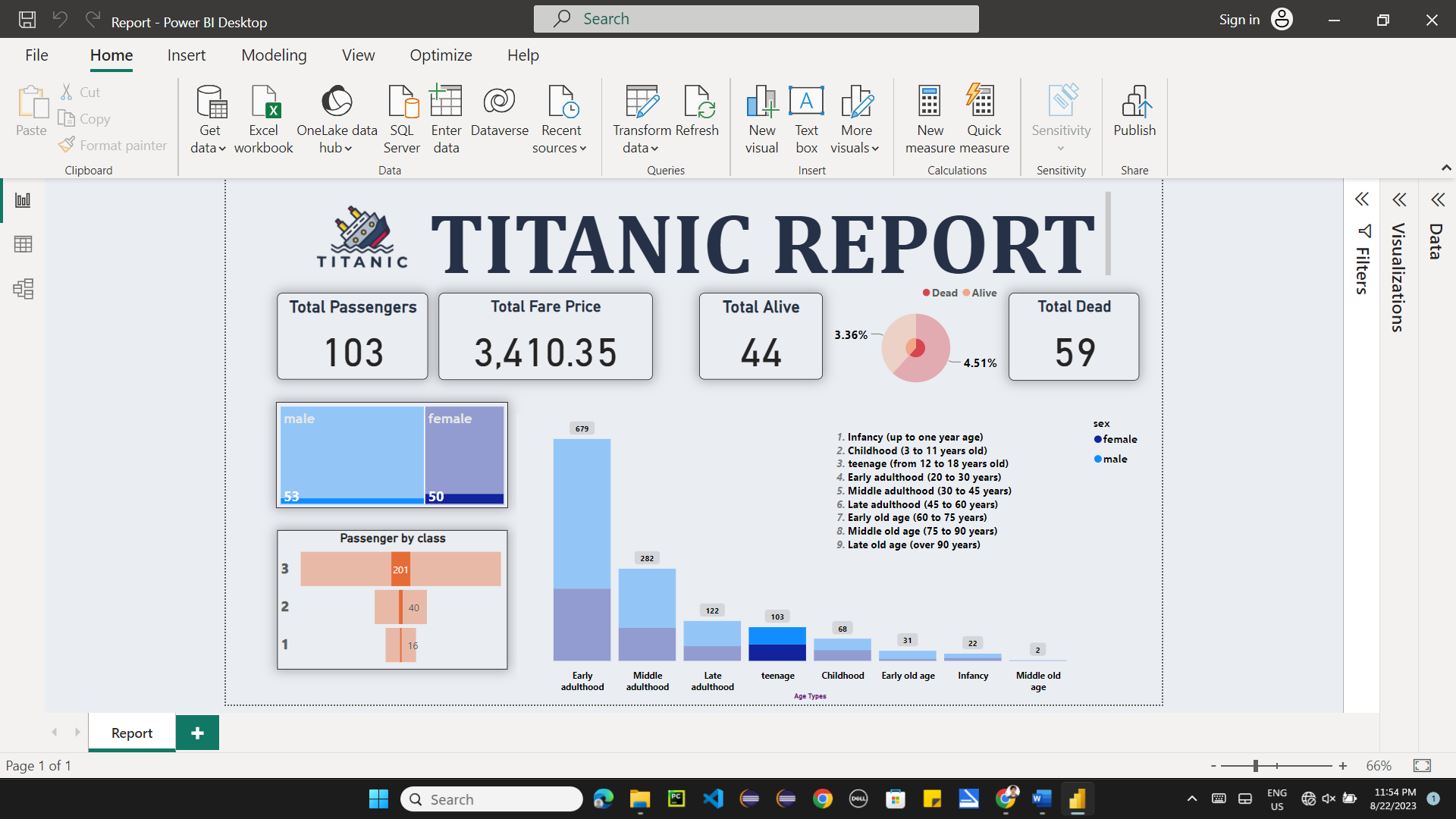
**Classification based on**

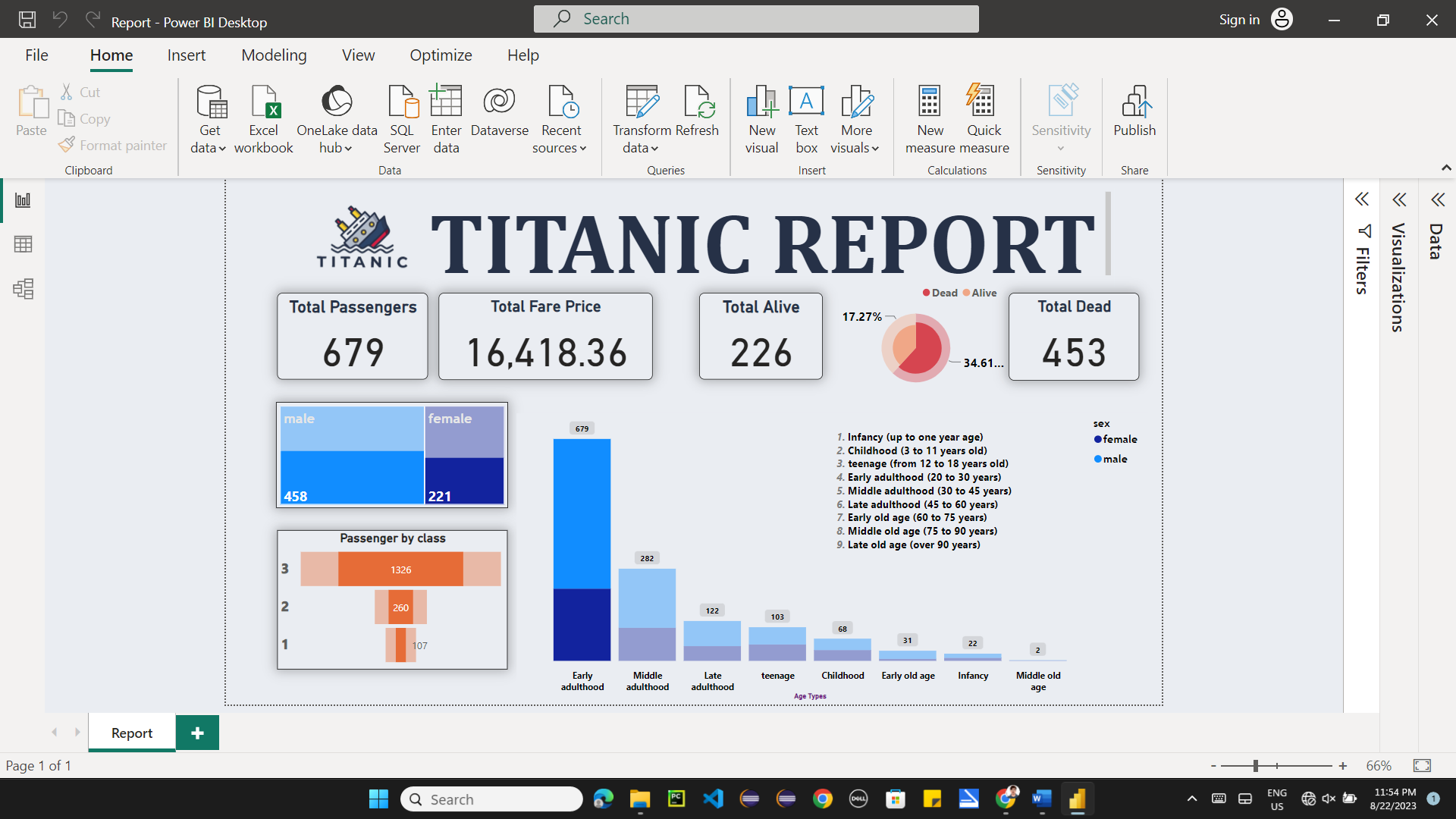
**Age Categories**

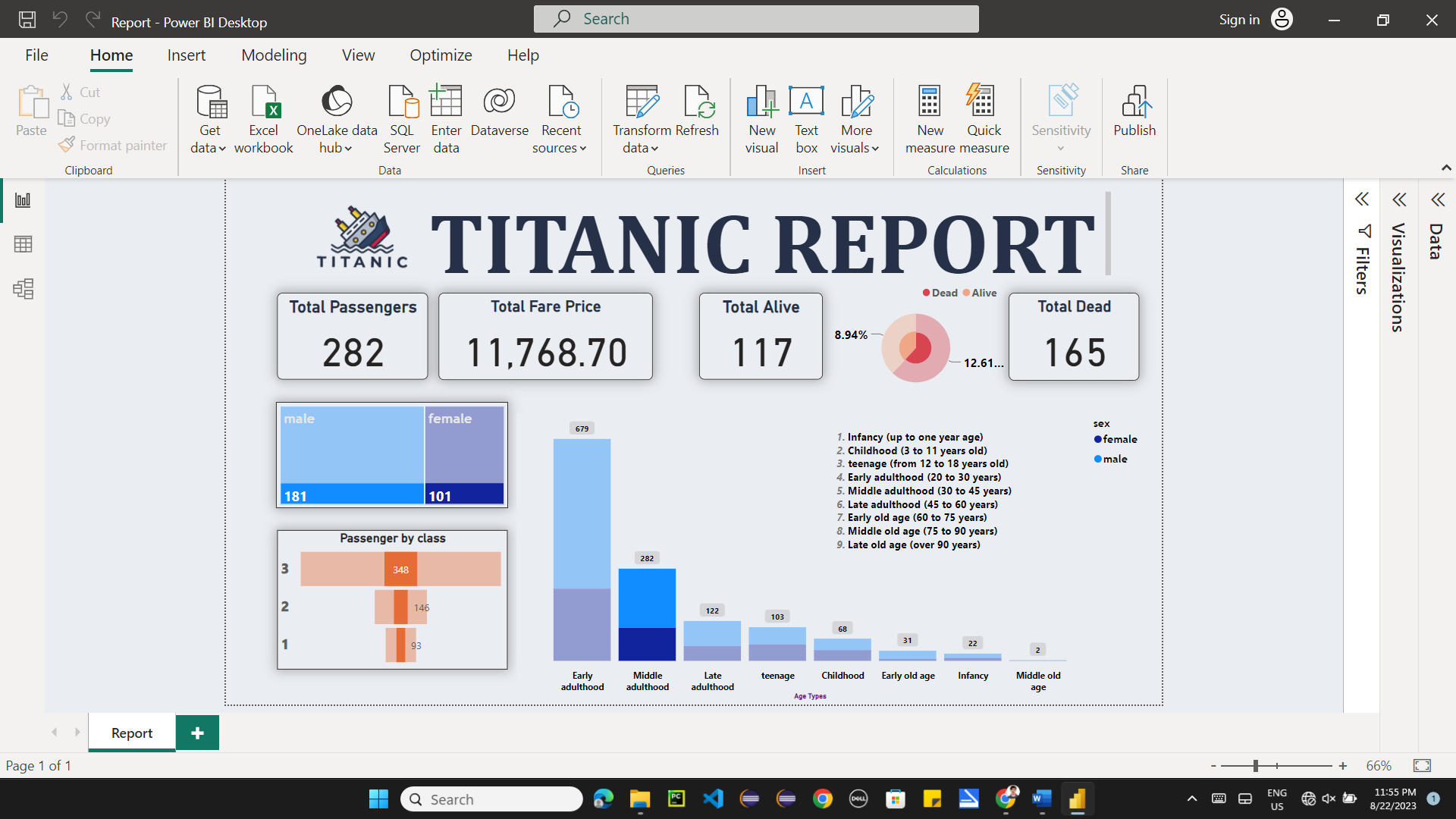
|  |
| --- |
| 1. **Infancy (neonate and up to one year age)** |
| 1. **Childhood (3 to 11 years old)** |
| 1. **Adolescence or teenage (from 12 to 18 years old)** |
| 1. **Early adulthood (20 to 30 years)** |
| 1. **Middle adulthood (30 to 45 years)** |
| 1. **Late adulthood (45 to 60 years)** |
| 1. **Early old age (60 to 75 years)** |
| 1. **Middle old age (75 to 90 years)** |
| 1. **Late old age (over 90 years)** |

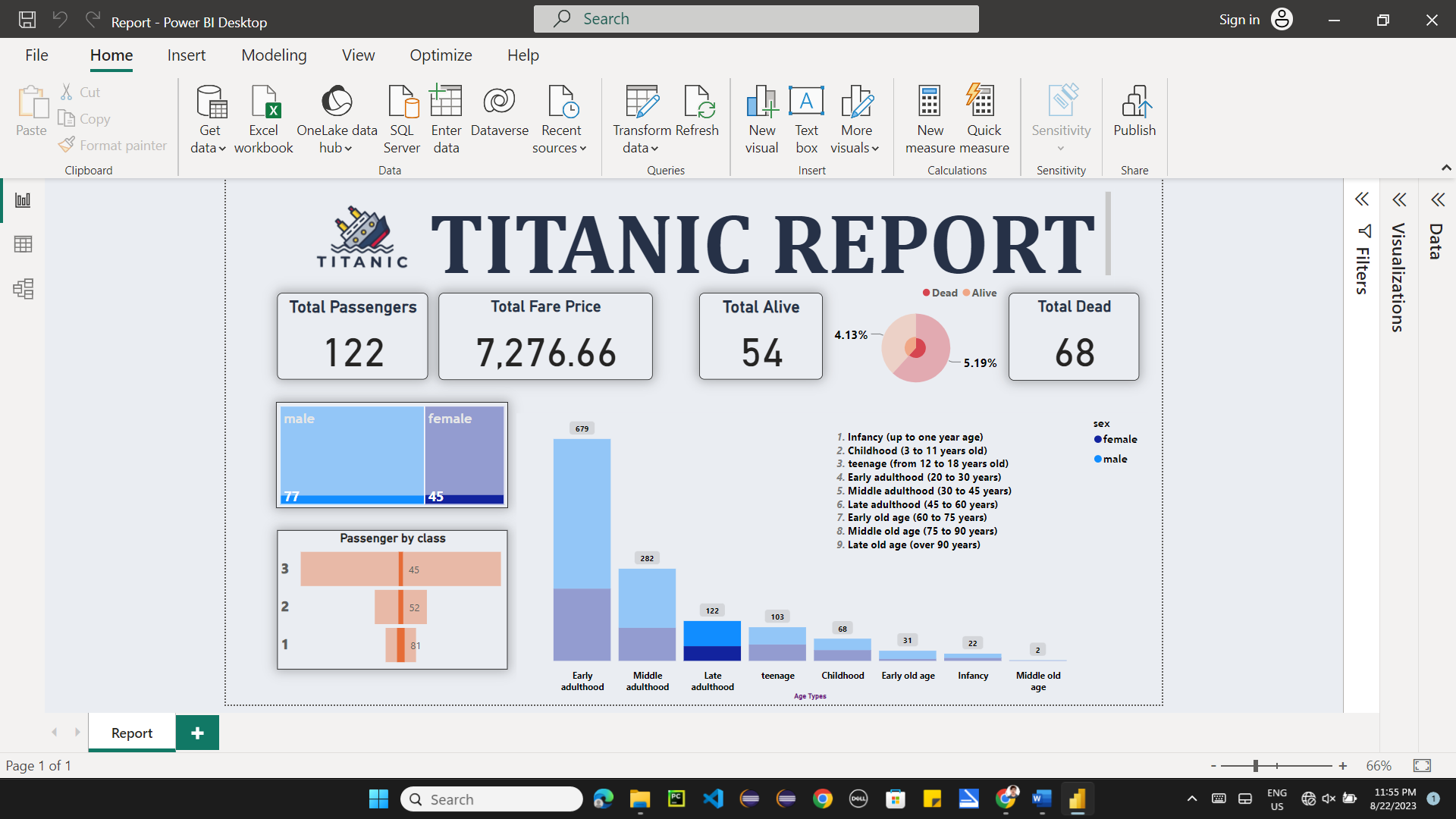


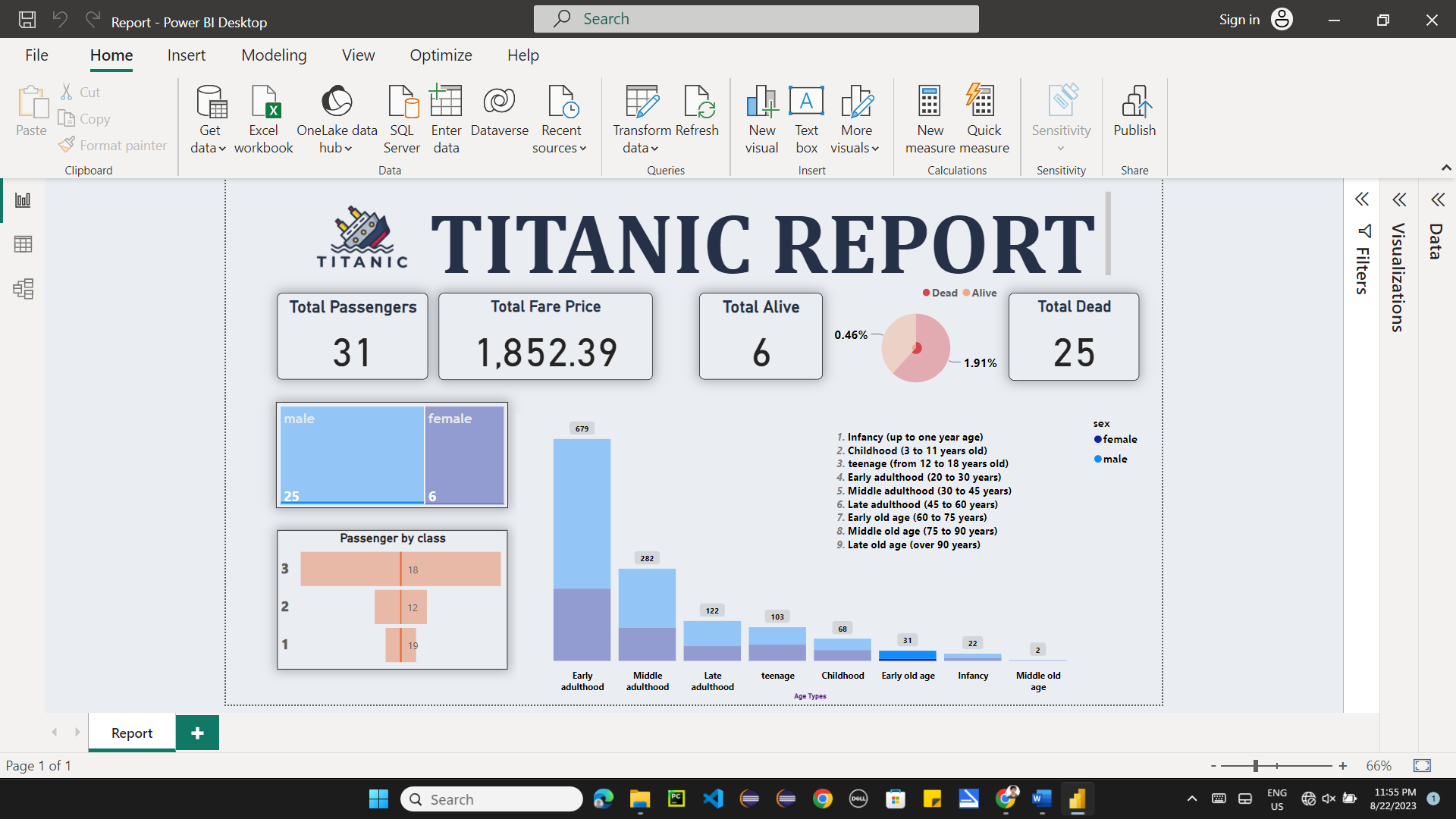


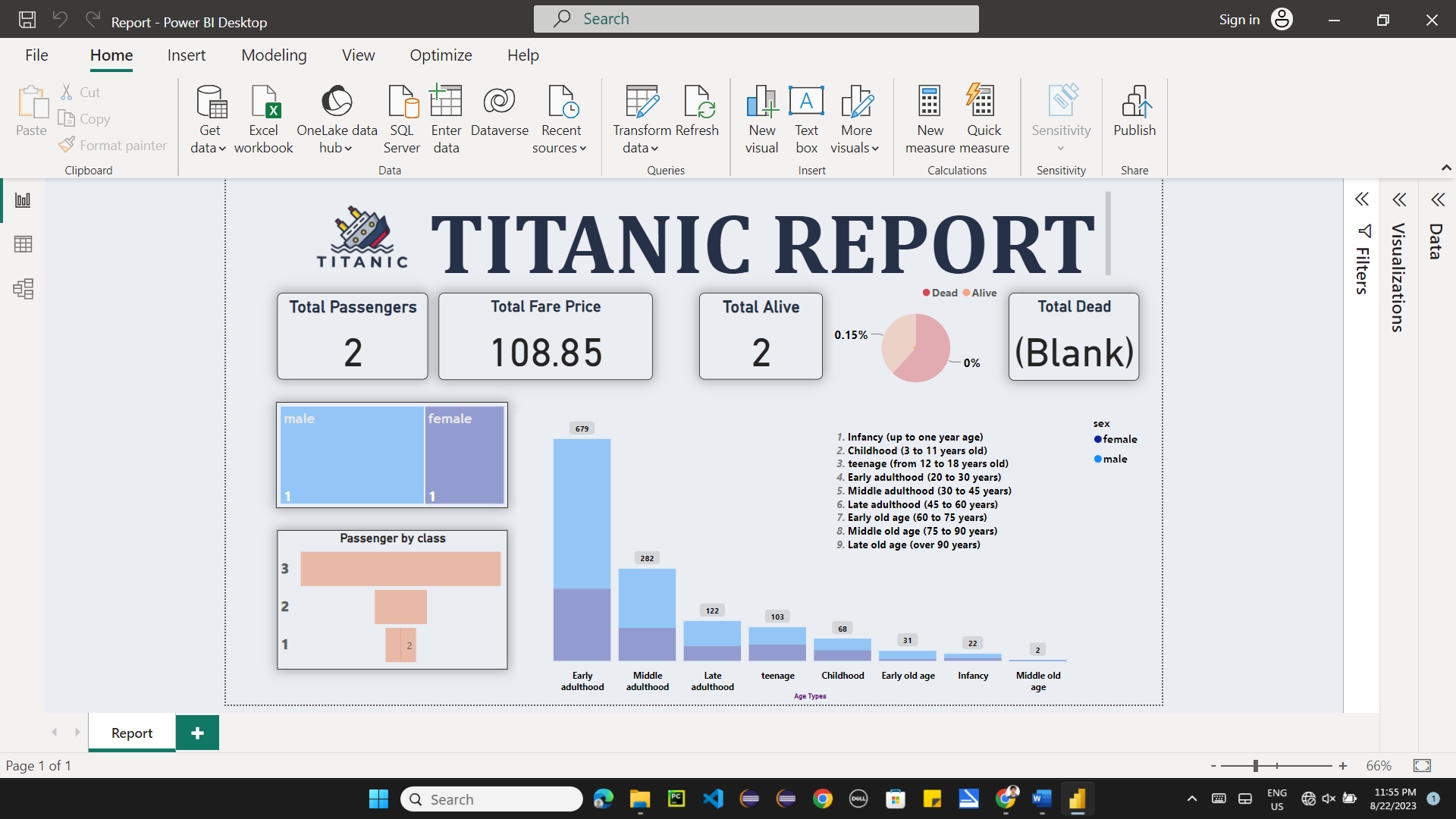












**Classification based on**

**Alive and Dead**

