INSIGHT FROM TITANIC MISHAP.

Analysis conducted

By

Ezuruike Chimere Enyinnaya

July, 2024

Outline

- **Executive Summary**
- **Introduction**
- **Methodology**
- **Analysis/interpretations**
- **visualizations**
- **Findings/implications**
- **Conclusion**
- **Appendix**

Executive Summary

- The Titanic dataset analysis provides profound insights into the tragic maritime disaster.
- Key findings reveal:- Women, children, and first-class passengers exhibited higher survival rates.
- Age, class, and gender significantly influenced survival outcomes.
- The cost of tickets for each passenger class impacted passenger numbers.
- Family size correlated directly with mortality rates.
- Infants comprised 4.2% of the ship's passengers.
- Alarmingly, over 60% of young people lost their lives.
- These findings offer valuable lessons for disaster preparedness, emergency response, and social responsibility.

Introduction

This analysis investigates the Titanic dataset to identify key variables influencing survival rates, shedding light on the complex interplay of factors that determined fate on that fateful night.

Methodology

- This analysis employed a quantitative approach using descriptive statistics and data visualization. The Titanic dataset was cleaned, transformed, and analyzed using Python with libraries Pandas, NumPy, and Matplotlib. Key variables were selected, and summary statistics, visualization, and inferential statistics were used to identify patterns and relationships.

Home | titanic project

localhost:8888/notebooks/OneDrive/Desktop/colt/DataAnalysis/data/titanic%20project.ipynb?

jupyter titanic project Last Checkpoint: 22 hours ago

File Edit View Run Kernel Settings Help Trusted

Code JupyterLab Python 3 (ipykernel)

[21]:
```
titanic.head(3)
```

[21]:

	pclass	survived	name	sex	age	sibsp	parch	ticket	fare	cabin	embarked	boat	body	home.dest
0	1	1	Allen, Miss. Elisabeth Walton	female	29.0000	0	0	24160	211.3375	B5	S	2	?	St Louis, MO
1	1	1	Allison, Master. Hudson Trevor	male	0.9167	1	2	113781	151.55	C22 C26	S	11	?	Montreal, PQ / Chesterville, ON
2	1	0	Allison, Miss. Helen Loraine	female	2.0000	1	2	113781	151.55	C22 C26	S	?	?	Montreal, PQ / Chesterville, ON

[ ]:
```
```

[11]:
```
titanic['age'] = titanic['age'].fillna(titanic['age'].median())
```

[9]:
```
titanic['age'] = pd.to_numeric(titanic['age'], errors='coerce')
```

Distribution Of Survivals And Non- Survivals

Home | titanic project

localhost:8888/notebooks/OneDrive/Desktop/colt/DataAnalysis/data/titanic%20project.ipynb?

jupyter titanic project Last Checkpoint: 22 hours ago

File Edit View Run Kernel Settings Help

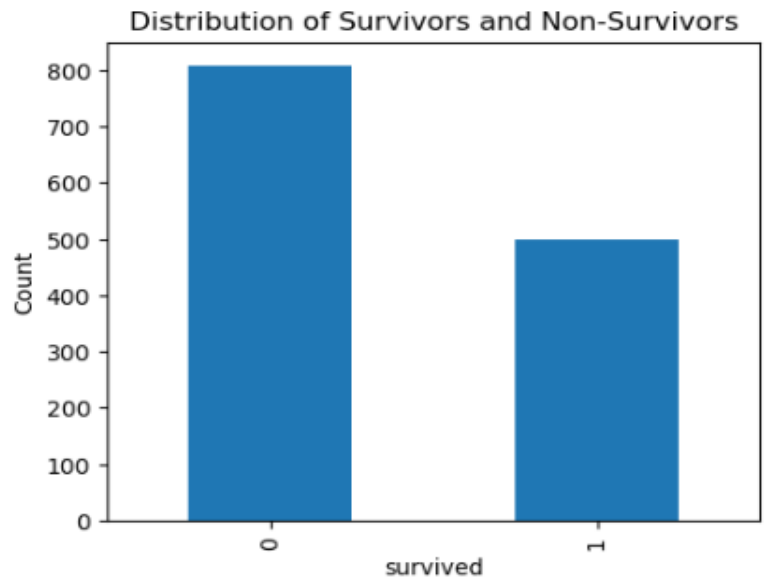
Trusted

Code

JupyterLab Python 3 (ipykernel)

[186]:
```
plt.figure(figsize=(5, 4))
titanic['survived'].value_counts().plot.bar()
plt.ylabel('Count')
plt.title('Distribution of Survivors and Non-Survivors')
```

[186]:
```
Text(0.5, 1.0, 'Distribution of Survivors and Non-Survivors')
```



[ ]:

[49]:
```
plt.figure(figsize=(8, 6))
plt.bar(titanic['survived'].value_counts().index, titanic['survived'].value_counts())
```

Findings:

- 1 and 0 in the chart above represents survial and death rate respectively
- There are over 700 deaths as compare to 500 survivals.

Survival By Gender

Home | titanic project

localhost:8888/notebooks/OneDrive/Desktop/colt/DataAnalysis/data/titanic%20project.ipynb?

jupyter titanic project Last Checkpoint: 22 hours ago

File Edit View Run Kernel Settings Help

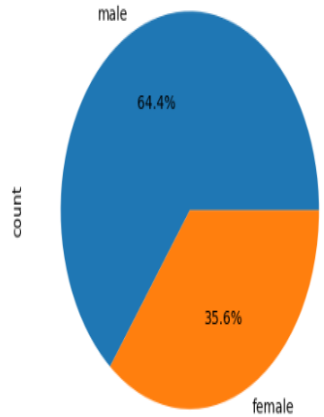
Trusted

Code

JupyterLab Python 3 (ipykernel)

```
[208]: titanic['sex'].value_counts().plot.pie(autopct="%1.1f%%")
```

```
[208]: <Axes: ylabel='count'>
```



Home | titanic project

localhost:8888/notebooks/OneDrive/Desktop/colt/DataAnalysis/data/titanic%20project.ipynb?

jupyter titanic project Last Checkpoint: 22 hours ago

File Edit View Run Kernel Settings Help

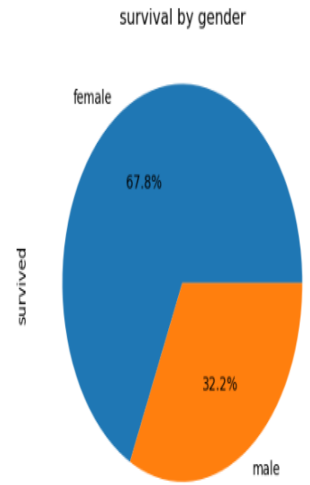
Trusted

Code

JupyterLab Python 3 (ipykernel)

```
[202]: titanic.groupby('sex')['survived'].sum().plot.pie(autopct="%1.1f%%")
       plt.title('survival by gender')
```

```
[202]: Text(0.5, 1.0, 'survival by gender')
```



Key Findings:

- There were 843 (64.4%)Males on board, only 161 (32.2%) survived.
- There were 466 (35.6%)females on board, about (67.5%) survived.
- This indicates that so many male travelers lost their lives on that fateful day.

% Of People In Each Passenger Class And Their Survival Rate

Home | titanic project

localhost:8888/notebooks/OneDrive/Desktop/colt/DataAnalysis/data/titanic%20project.ipynb?

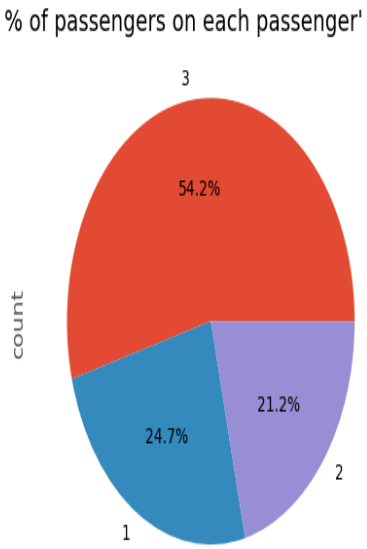
jupyter titanic project Last Checkpoint: 23 hours ago

File Edit View Run Kernel Settings Help | Trusted

Code | JupyterLab | Python 3 (ipykernel)

```
Name: count, dtype: int64
```

[251]:
```
titanic.pclass.value_counts().plot.pie(autopct="%1.1f%%")
plt.title("% of passengers on each passenger' class")
plt.show()
```



Home | titanic project

localhost:8888/notebooks/OneDrive/Desktop/colt/DataAnalysis/data/titanic%20project.ipynb?

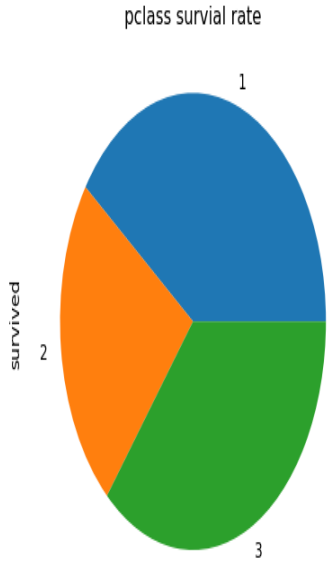
jupyter titanic project Last Checkpoint: 23 hours ago

File Edit View Run Kernel Settings Help | Trusted

Code | JupyterLab | Python 3 (ipykernel)

[216]:
```
titanic.groupby('pclass')['survived'].sum().plot.pie()
plt.title('pclass survial rate')
```

[216]:
```
Text(0.5, 1.0, 'pclass survial rate')
```



Passengers Class vs Fare

Home | titanic project

localhost:8888/notebooks/OneDrive/Desktop/colt/DataAnalysis/data/titanic%20project.ipynb?

jupyter titanic project Last Checkpoint: 23 hours ago

File Edit View Run Kernel Settings Help

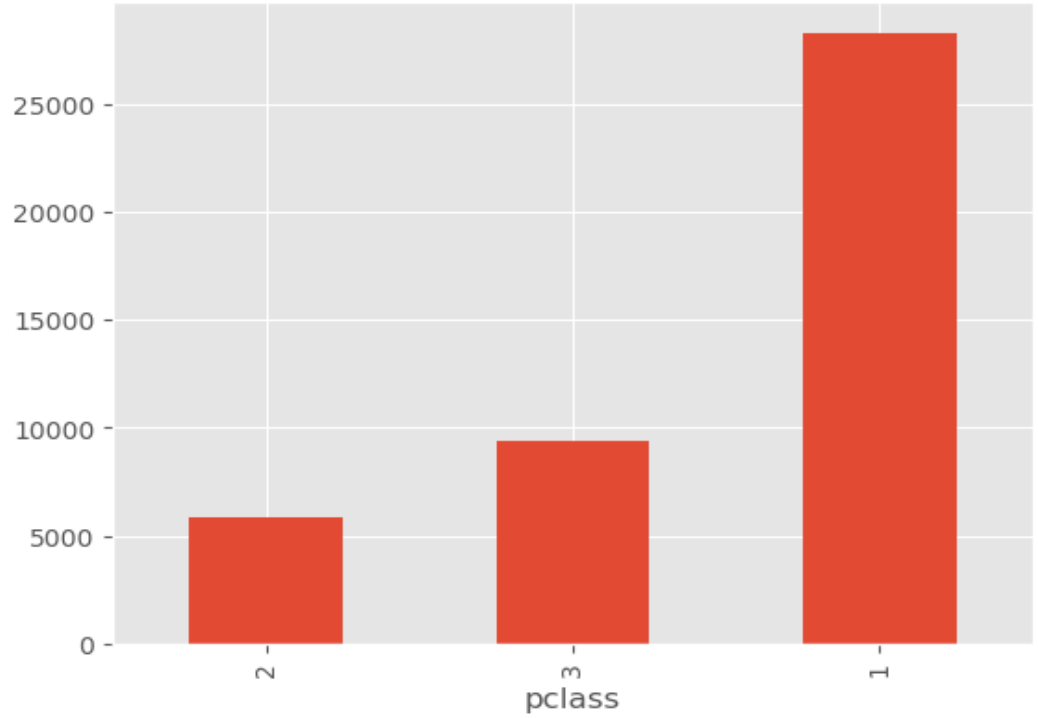
Trusted

Code

JupyterLab Python 3 (ipykernel)

```
Name: count, dtype: int64
```

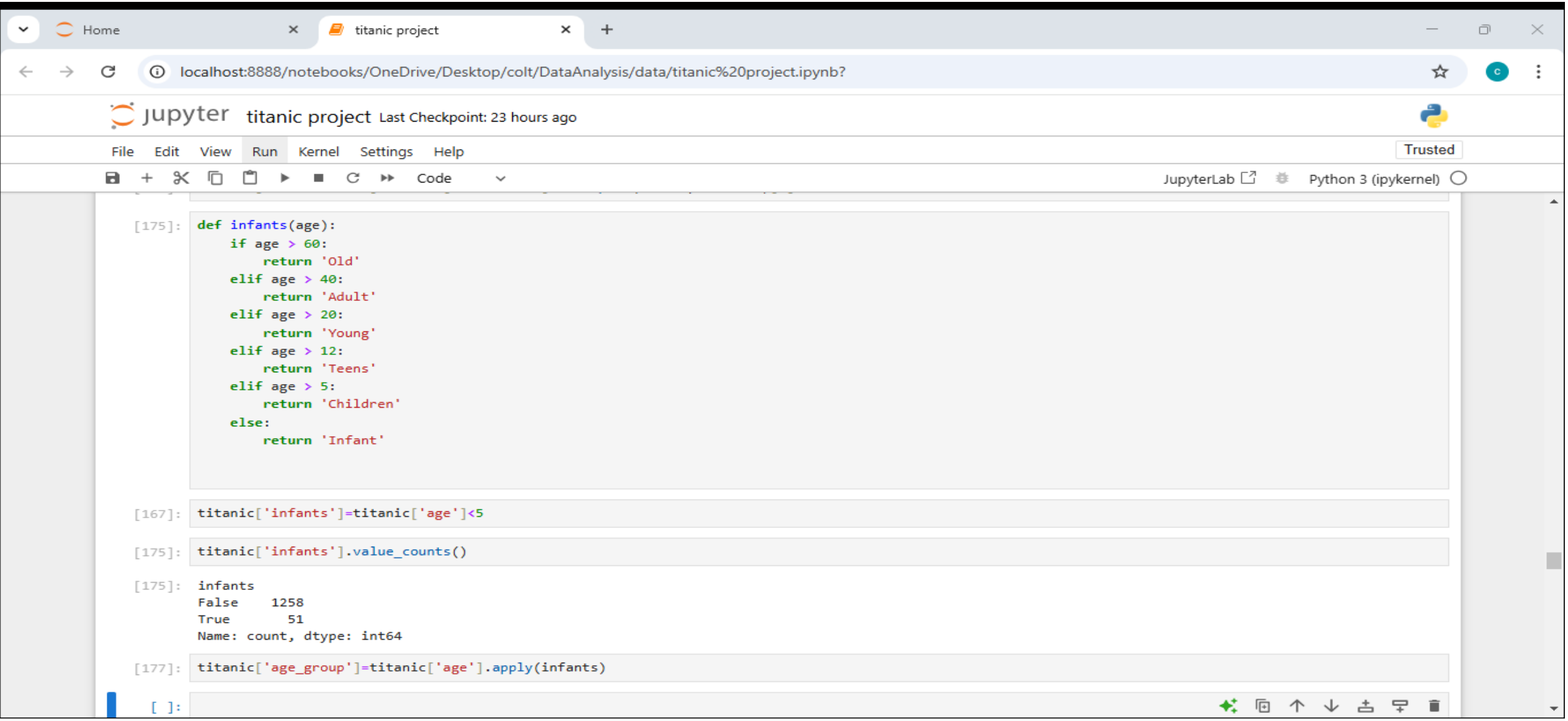
[249]:
```
titanic.groupby('pclass')['fare'].sum().sort_values().plot.bar()
# plt.yticks([1, 2, 3], ['Ist_Class', '2nd_Class', '3rd_class'])
plt.show()
```



Findings

- There were large number of passengers in third class as well as more death of passengers.
- First class passengers seams safer though with fewer passengers on board. This is because the average fare is above $80 compare to around $20 for second class and less than $15 for third class. And this disparity influence the rush to 3rd class which lead to loss of many lives.

Age group analysis



Age Group Analysis 2

Home | titanic project

localhost:8888/notebooks/OneDrive/Desktop/colt/DataAnalysis/data/titanic%20project.ipynb?

jupyter titanic project Last Checkpoint: 23 hours ago

File Edit View Run Kernel Settings Help

Trusted

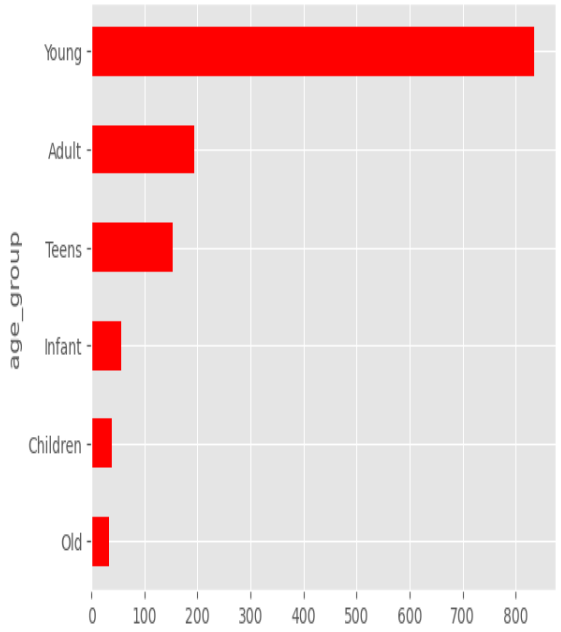
Code

JupyterLab Python 3 (ipykernel)

```
[ ]:
```

```
[247]: titanic['age_group'].value_counts().sort_values().plot(kind='barh',color='red')
```

```
[247]: <Axes: ylabel='age_group'>
```



Home | titanic project

localhost:8888/notebooks/OneDrive/Desktop/colt/DataAnalysis/data/titanic%20project.ipynb?

jupyter titanic project Last Checkpoint: 23 hours ago

File Edit View Run Kernel Settings Help

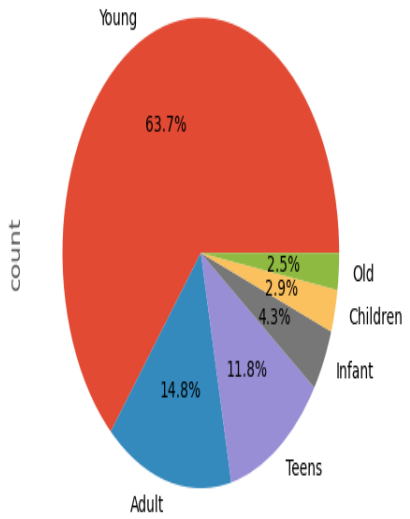
Trusted

Code

JupyterLab Python 3 (ipykernel)

```
[237]: titanic['age_group'].value_counts().plot.pie(autopct="%1.1f%%")
```

```
[237]: <Axes: ylabel='count'>
```




```
[227]: titanic['age_group'].value_counts()
```

```
[227]: age_group
       Young      834
```

Age Distribution(histogram)

Home | titanic project

localhost:8888/notebooks/OneDrive/Desktop/colt/DataAnalysis/data/titanic%20project.ipynb?

jupyter titanic project Last Checkpoint: 23 hours ago

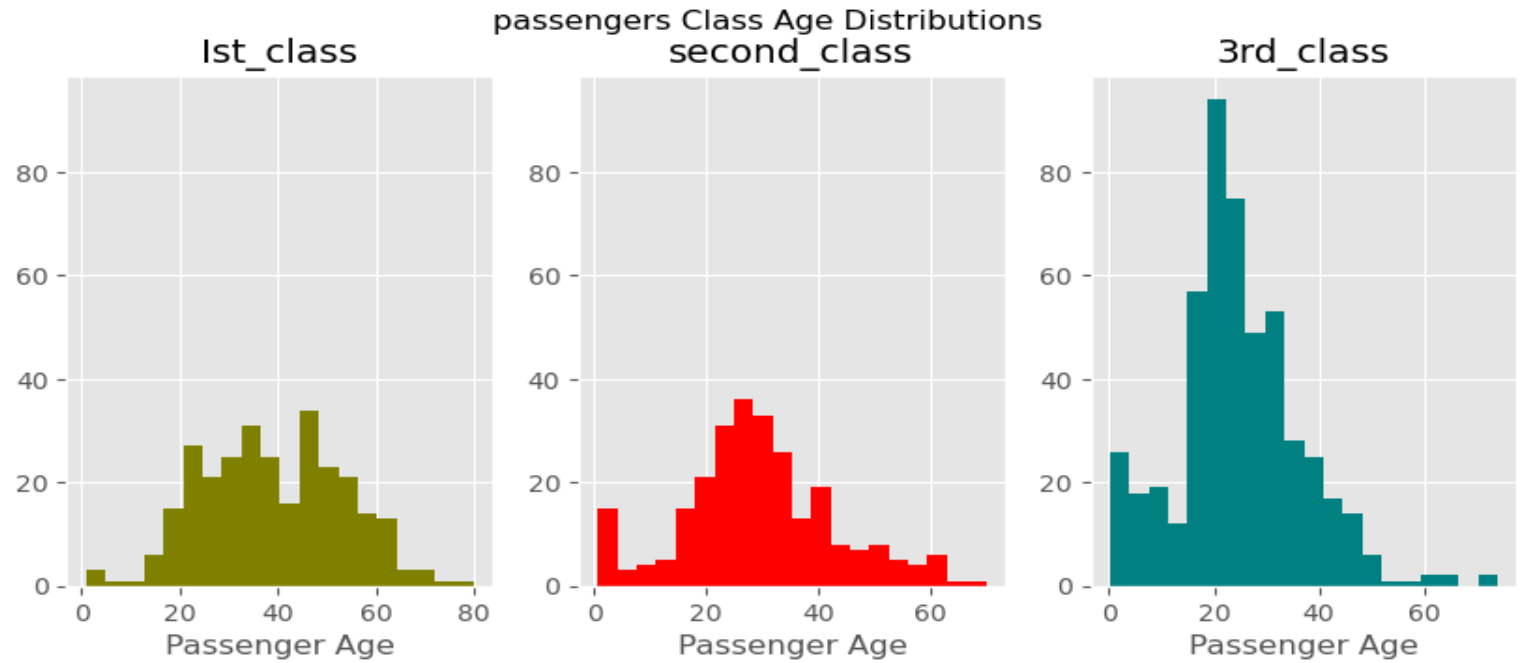
File Edit View Run Kernel Settings Help

Trusted

Code

JupyterLab | Python 3 (ipykernel)

```
plt.suptitle('passengers Class Age Distributions')
plt.show()
```



[149]:
```
titanic.groupby('pclass')['survived'].sum().reset_index()
```

Age distribution code

Home | titanic project

localhost:8888/notebooks/OneDrive/Desktop/colt/DataAnalysis/data/titanic%20project.ipynb?

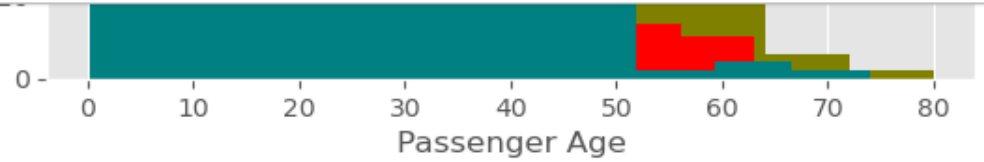
jupyter titanic project Last Checkpoint: 23 hours ago

File Edit View Run Kernel Settings Help

Trusted

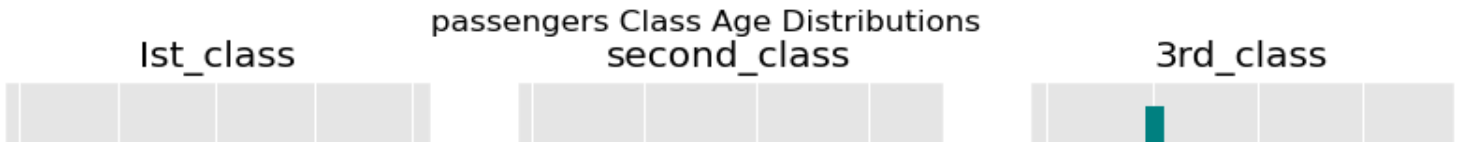
Code

JupyterLab Python 3 (ipykernel)



[115]:

```
plt.figure(figsize=(10,4))
plt.subplot(1,3,1)
ax=plt.subplot(1,3,1)
plt.title('Ist_class')
plt.hist(first_class['age'],color='olive',label='first_class',bins=20)
plt.xlabel('Passenger Age')
plt.subplot(1,3,2,sharey=ax)
plt.title('second_class')
plt.hist(second_class['age'],color='red',label='second_class',bins=20)
plt.xlabel('Passenger Age')
plt.subplot(1,3,3,sharey=ax)
plt.title('3rd_class')
plt.hist(third_class['age'],color='teal',label='third_class',bins=20)
plt.xlabel('Passenger Age')
plt.suptitle('passengers Class Age Distributions')
plt.show()
```



Family size Analysis

Home | titanic project

localhost:8888/notebooks/OneDrive/Desktop/colt/DataAnalysis/data/titanic%20project.ipynb?

jupyter titanic project Last Checkpoint: yesterday

File Edit View Run Kernel Settings Help

Trusted

Code

JupyterLab Python 3 (ipykernel)

[67]:
```
titanic['fam_size']=titanic['sibsp']+titanic['parch']
titanic['fam_size'].min()
```

[67]:
```
0
```

[75]:
```
def fam_sizes(row):
    if row['fam_size'] < 1:
        return 'solo_traveller'
    elif row['fam_size'] < 4:
        return 'Average_family'
    else:
        return 'large_family'
```

[77]:
```
titanic['family_size_category'] = titanic.apply(fam_sizes, axis=1)
```

[87]:
```
titanic['family_size_category'].value_counts()
```

[87]:
```
family_size_category
solo_traveller    790
Average_family    437
large_family       82
Name: count, dtype: int64
```

[85]:
```
titanic['family_size_category'].value_counts().plot.pie(autopct="%1.1f%%")
```

Family size analysis2

Home | titanic project

localhost:8888/notebooks/OneDrive/Desktop/colt/DataAnalysis/data/titanic%20project.ipynb?

jupyter titanic project Last Checkpoint: yesterday

File Edit View Run Kernel Settings Help

Trusted

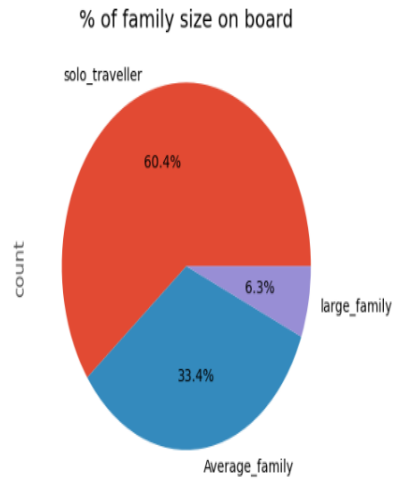
Code

JupyterLab Python 3 (ipykernel)

```
large_family      82
Name: count, dtype: int64
```

[261]:
```
titanic['family_size_category'].value_counts().plot.pie(autopct="%1.1f%%")
plt.title('% of family size on board')
```

[261]: Text(0.5, 1.0, '% of family size on board')



[93]:
```
titanic.groupby('family_size_category')['survived'].mean()
```

[93]: family_size_category

Home | titanic project

localhost:8888/notebooks/OneDrive/Desktop/colt/DataAnalysis/data/titanic%20project.ipynb?

jupyter titanic project Last Checkpoint: yesterday

File Edit View Run Kernel Settings Help

Trusted

Code

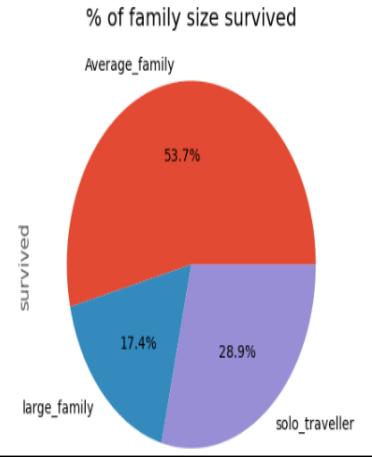
JupyterLab Python 3 (ipykernel)

[93]:
```
titanic.groupby('family_size_category')['survived'].mean()
```

[93]:
```
family_size_category
Average_family    0.562929
large_family      0.182927
solo_traveller    0.302532
Name: survived, dtype: float64
```

[259]:
```
titanic.groupby('family_size_category')['survived'].mean().plot.pie(autopct="%1.1f%%")
plt.title('% of family size survived')
```

[259]: Text(0.5, 1.0, '% of family size survived')



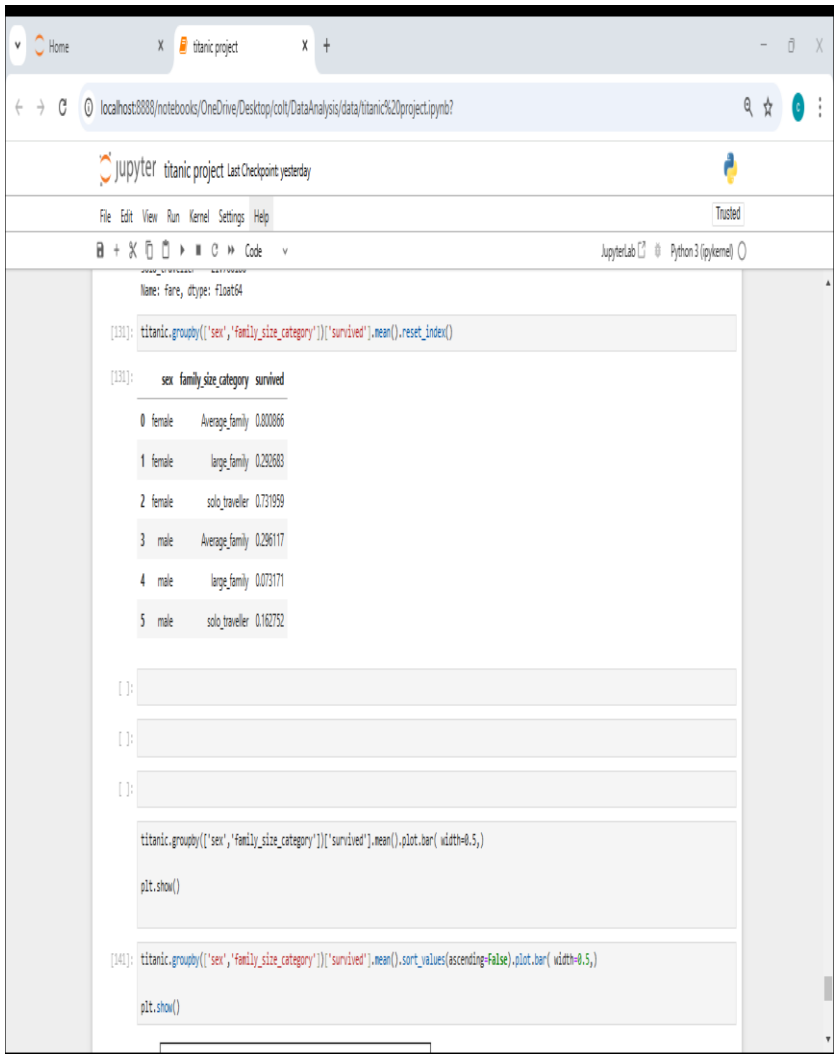
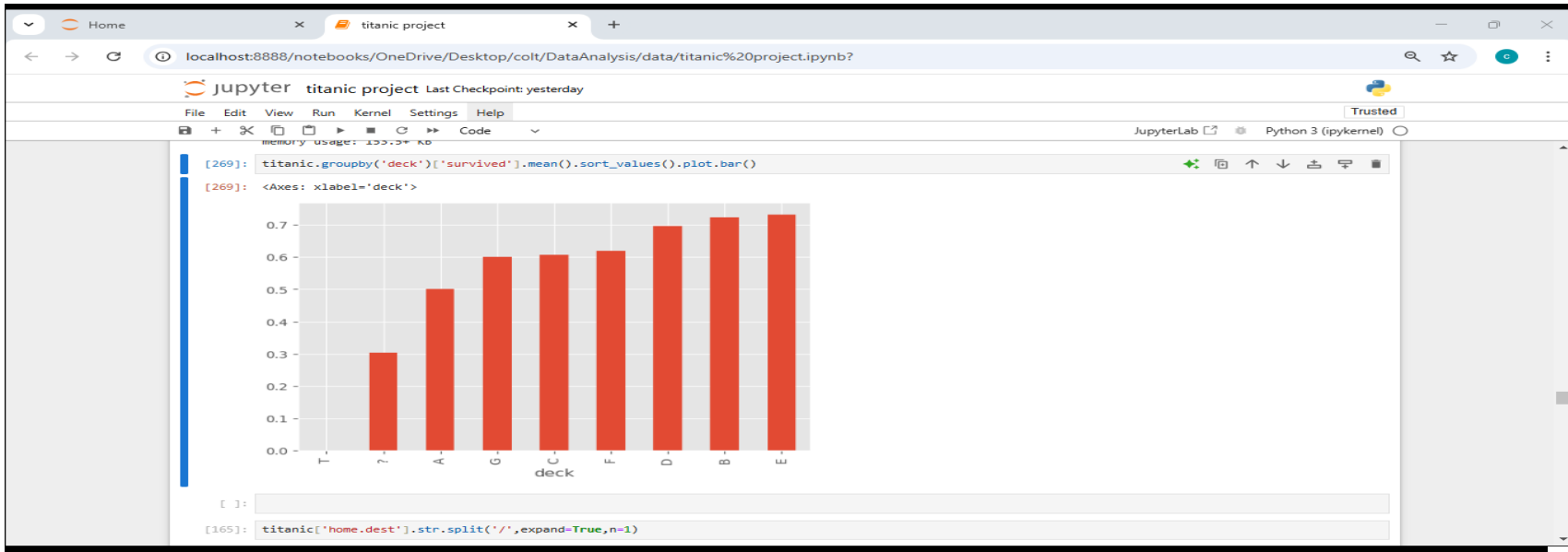
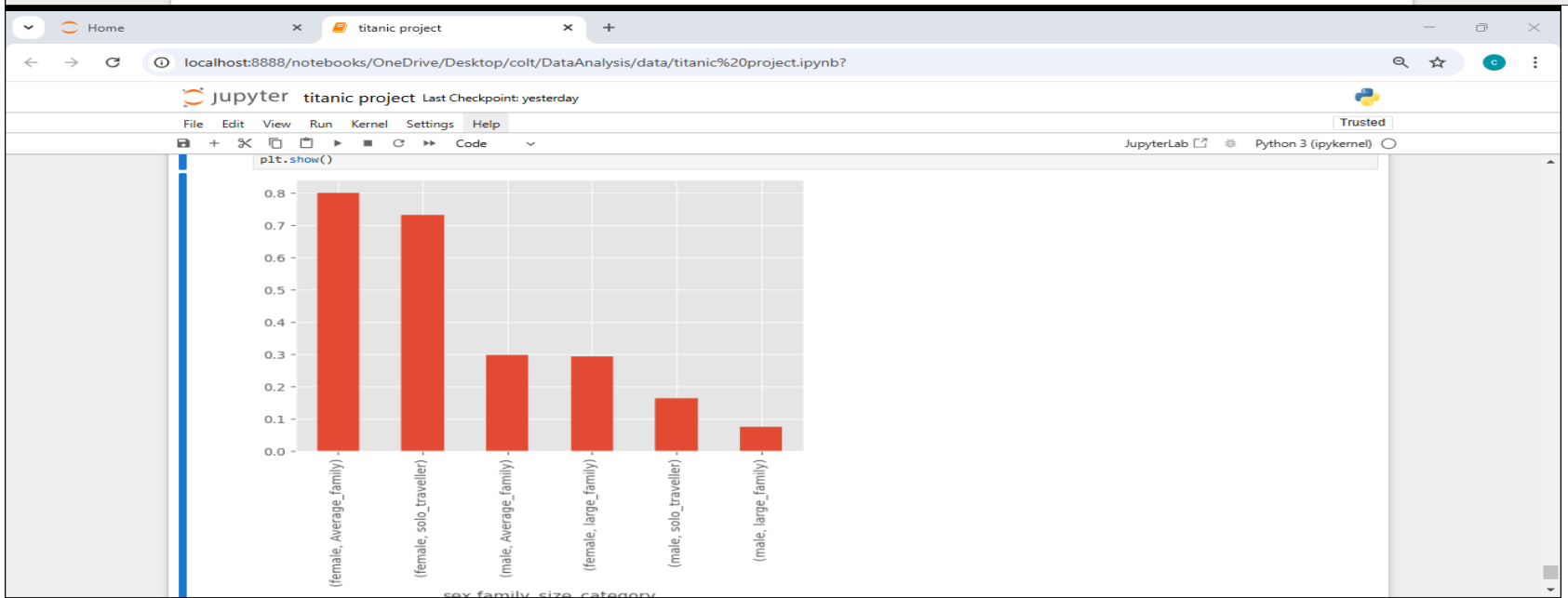
Findings

- There were so many young persons on board and so many lost their lives as well.
- Very few old persons were involved in the disaster
- Also a good number of infants and children lost their lives
- Average family size survived more than solo-travelers and larger family size respectively.

Overall Findings

- **Demographic Disparities**: The passenger population was predominantly male, The age distribution revealed a youthful demographic, with a majority of passengers between the ages of 20-40.
- **Socio-Economic Divide:** The analysis of passenger fares and classes exposed striking disparities, with first-class passengers paying substantially higher fares than their second- and third-class counterparts.
- **Age and Survival:** Young people, despite being more numerous, suffered significant losses, with over 60% of young passengers losing their lives.
- **Family and Survival:** Family size correlated directly with mortality rates, indicating that larger families faced higher risks.

Conclusion/Recommendation

- Indeed one's socio-economic class goes a long way in determining his/her survival in life.
- This analysis proved that the imbalance in socio economic status of the titanic passengers lead to the loss of many lives especially the poor.
- it also shows that the organizers of this adventure did not make adequate provision for life jacket especially for those in third class category.
- The result also shows that young people like adventures more than the old.
- **Recommendation:**
- **Improved Emergency Preparedness:** Develop and implement more effective emergency response plans, including regular drills and training for crew members.
- **Inclusive and Equitable Policies:** Develop policies that prioritize the safety and well-being of vulnerable populations, including women, children, and passengers from lower socio-economic backgrounds.
- I recommend that only few passengers should make the journey and should travel light as well.

Appendix (Data source: Kaggle)