

titanic

January 5, 2026

Titatnic Dataset

Section	Description
1. Problem Framing	Research questions, formal hypotheses (H_0/H_1), significance levels
2. Raw Data Inspection	Data loading, structure analysis, column descriptions
3. Missingness Analysis	MCAR/MAR/MNAR detection, missingness patterns, data leakage checks
4. Descriptive Statistics	Stratified analysis by gender, class, age, family size
5. Hypothesis Testing	Chi-square, t-tests (two-tailed & one-tailed), ANOVA with post-hoc
6. Correlation Analysis	Pearson & Spearman correlations, effect sizes, R^2 interpretation
7. Feature Engineering	FamilySize, IsAlone, FarePerPerson, Title extraction, HasCabin
8. Anomaly Detection	IQR & Z-score methods, outlier analysis for Fare, Age, FamilySize
9. Missing Data Handling	Conditional imputation by Title/Pclass, domain-informed decisions
10. Visualization	12-panel dashboard, statistical storytelling narrative
11. Bias & Ethics	Confounding, selection bias, causation vs correlation, ethical considerations
12. Executive Summary	Key findings, confidence table, recommendations, next steps

Section	Description
1. Problem Framing	Research questions, formal hypotheses (H0/H1), significance levels
2. Raw Data Inspection	Data loading, structure analysis, column descriptions
3. Missingness Analysis	MCAR/MAR/MNAR detection, missingness patterns, data leakage checks
4. Descriptive Statistics	Stratified analysis by gender, class, age, family size
5. Hypothesis Testing	Chi-square, t-tests (two-tailed & one-tailed), ANOVA with post-hoc
6. Correlation Analysis	Pearson & Spearman correlations, effect sizes, R ² interpretation
7. Feature Engineering	FamilySize, IsAlone, FarePerPerson, Title extraction, HasCabin
8. Anomaly Detection	IQR & Z-score methods, outlier analysis for Fare, Age, FamilySize
9. Missing Data Handling	Conditional imputation by Title/Pclass, domain-informed decisions
10. Visualization	12-panel dashboard, statistical storytelling narrative
11. Bias & Ethics	Confounding, selection bias, causation vs correlation, ethical considerations
12. Executive Summary	Key findings, confidence table, recommendations, next steps

```
[1]: import pandas as pd
import numpy as np
import sklearn
import seaborn as sn
import matplotlib as plt
```

Matplotlib is building the font cache; this may take a moment.

```
[2]: data = pd.read_csv("Data/titanic.csv")
data.head()
```

```
[2]:   PassengerId  Survived  Pclass \
0            1         0      3
1            2         1      1
2            3         1      3
3            4         1      1
4            5         0      3

                                                Name     Sex   Age  SibSp \
0          Braund, Mr. Owen Harris    male  22.0      1
1  Cumings, Mrs. John Bradley (Florence Briggs Th...  female  38.0      1
2                Heikkinen, Miss. Laina  female  26.0      0
3        Futrelle, Mrs. Jacques Heath (Lily May Peel)  female  35.0      1
4            Allen, Mr. William Henry    male  35.0      0

Parch      Ticket      Fare Cabin Embarked

```

0	0	A/5	21171	7.2500	NaN	S
1	0	PC	17599	71.2833	C85	C
2	0	STON/02.	3101282	7.9250	NaN	S
3	0		113803	53.1000	C123	S
4	0		373450	8.0500	NaN	S