2025/3/27, 20:15 CSV Data Insight

# CSV Insight Generator + AI Summary

Upload a .csv file to view descriptive stats, visualizations, and get LLM-generated insights.



Upload CSV File



Drag and drop file here

Limit 200MB per file • CSV

Browse files

X



train.csv 0.8MB



### Basic Info

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 8693 entries, 0 to 8692

Data columns (total 14 columns):

# Column Non-Null Count Dtype

- 0 PassengerId 8693 non-null object
- 1 HomePlanet 8492 non-null object
- 2 CryoSleep 8476 non-null object
- 3 Cabin 8494 non-null object
- 4 Destination 8511 non-null object
- 8514 non-null float64 5 Age
- 6 VIP 8490 non-null object
- 7 RoomService 8512 non-null float64
- 8 FoodCourt 8510 non-null float64
- 9 ShoppingMall 8485 non-null float64
- 8510 non-null float64 10 Spa

http://localhost:8501/ Page 1 of 14

11 VRDeck 8505 non-null float64 12 Name 8493 non-null object 13 Transported 8693 non-null bool dtypes: bool(1), float64(6), object(7)

memory usage: 891.5+ KB

# Summary Statistics

|        | PassengerId | HomePlanet | CryoSleep | Cabin  | Destination | Age     | VIP   | RoomService | FoodCo |
|--------|-------------|------------|-----------|--------|-------------|---------|-------|-------------|--------|
| count  | 8693        | 8492       | 8476      | 8494   | 8511        | 8514    | 8490  | 8512        | 8.     |
| unique | 8693        | 3          | 2         | 6560   | 3           | None    | 2     | None        | No     |
| top    | 0001_01     | Earth      | False     | G/734/ | TRAPPIST-1e | None    | False | None        | No     |
| freq   | 1           | 4602       | 5439      | 8      | 5915        | None    | 8291  | None        | No     |
| mean   | None        | None       | None      | None   | None        | 28.8279 | None  | 224.6876    | 458.0  |
| std    | None        | None       | None      | None   | None        | 14.489  | None  | 666.7177    | 1611.4 |
| min    | None        | None       | None      | None   | None        | 0       | None  | 0           |        |
| 25%    | None        | None       | None      | None   | None        | 19      | None  | 0           |        |
| 50%    | None        | None       | None      | None   | None        | 27      | None  | 0           |        |
| 75%    | None        | None       | None      | None   | None        | 38      | None  | 47          |        |
|        |             |            |           |        |             |         |       |             |        |

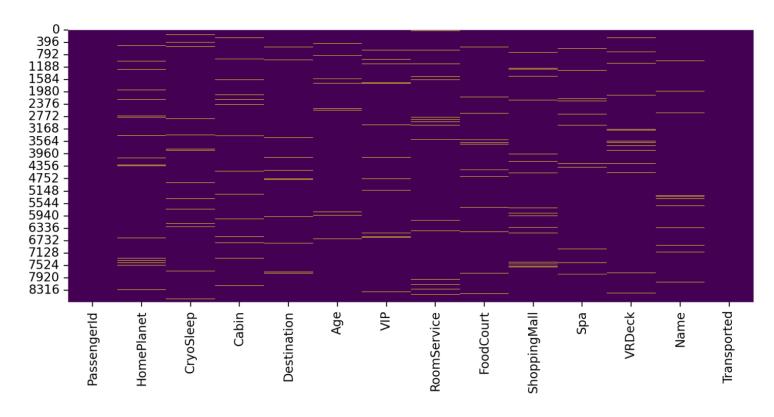
# Missing Values Table

|             | 0   |
|-------------|-----|
| HomePlanet  | 201 |
| CryoSleep   | 217 |
| Cabin       | 199 |
| Destination | 182 |
| Age         | 179 |

http://localhost:8501/ Page 2 of 14

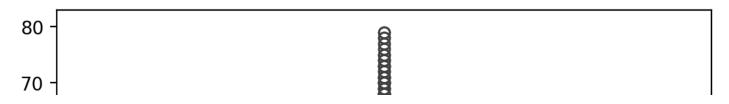
| VIP          | 203 |
|--------------|-----|
| RoomService  | 181 |
| FoodCourt    | 183 |
| ShoppingMall | 208 |
| Spa          | 183 |
|              |     |

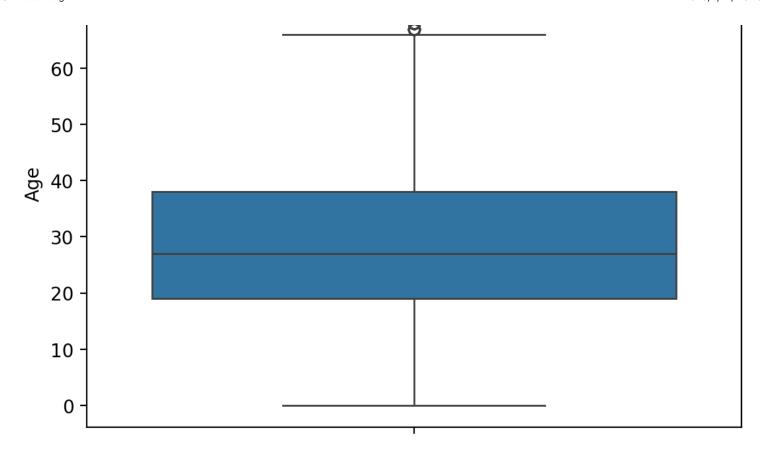
# Missing Values Heatmap



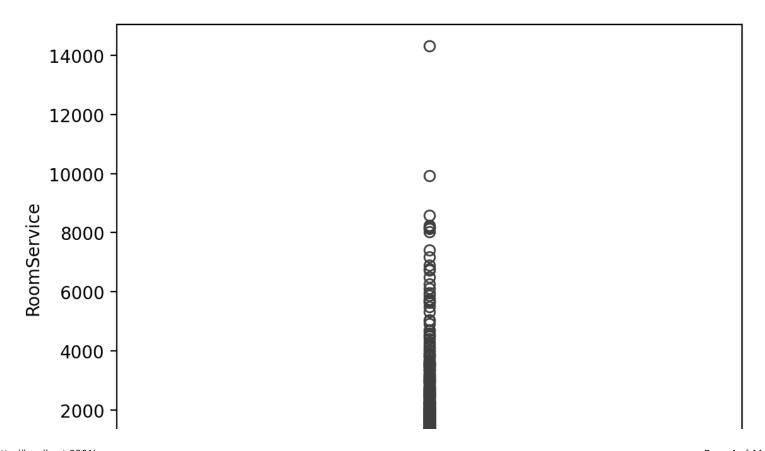
### **■** Distribution Plots

### Age Distribution





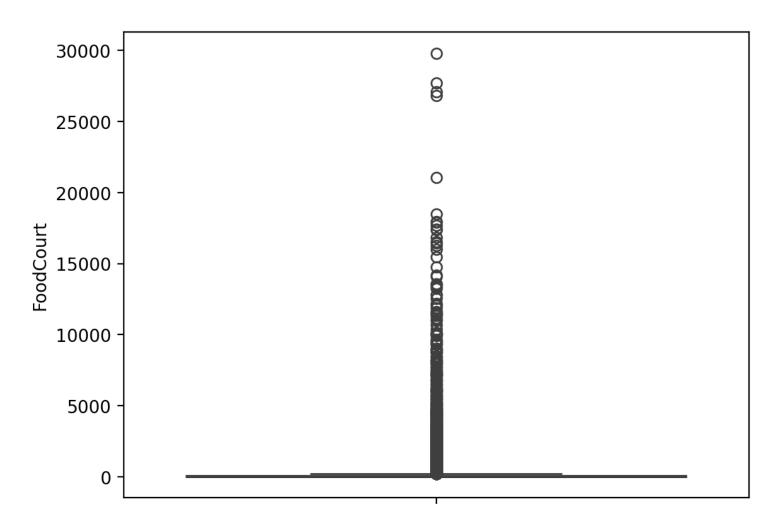
### RoomService Distribution



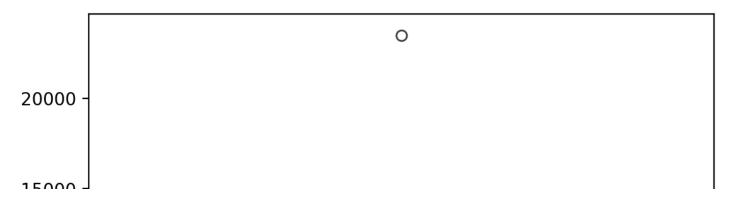
http://localhost:8501/ Page 4 of 14

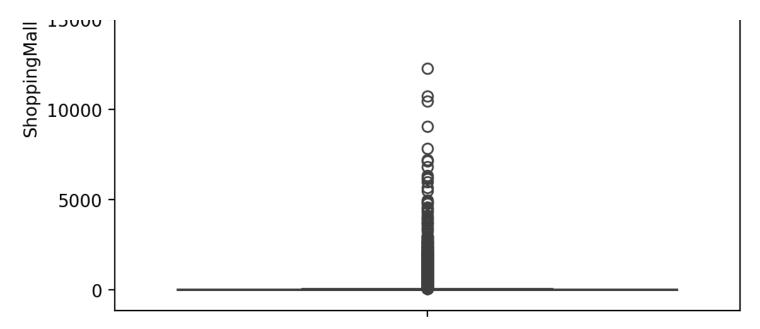


### FoodCourt Distribution

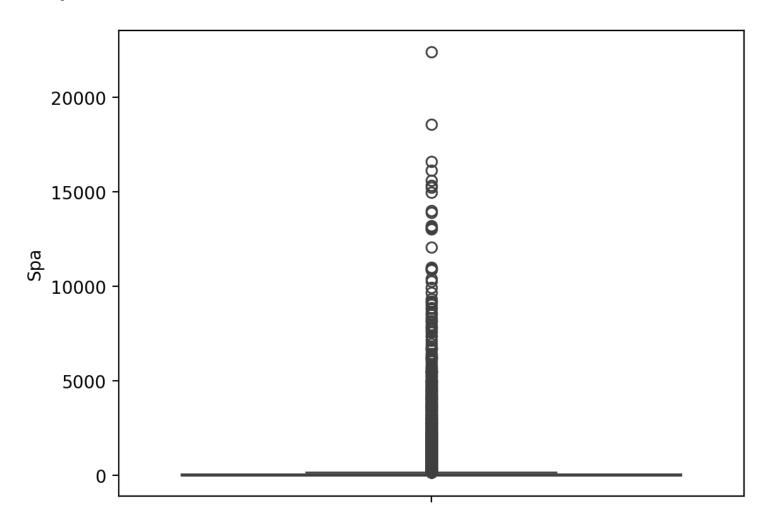


## ShoppingMall Distribution

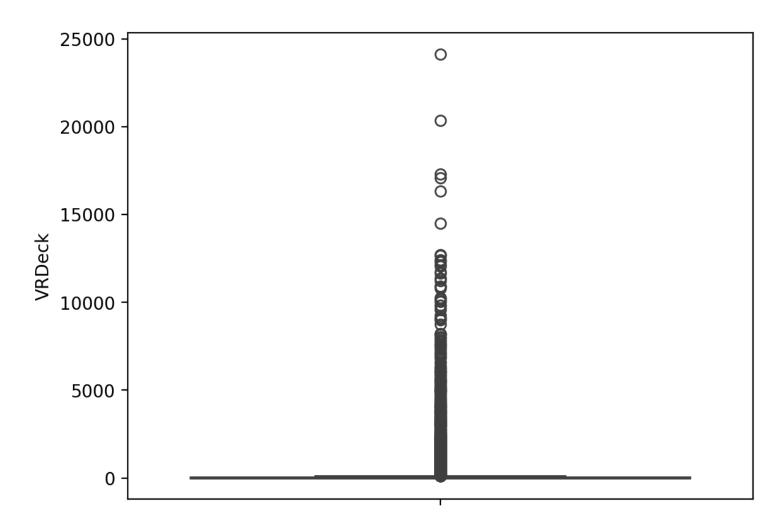




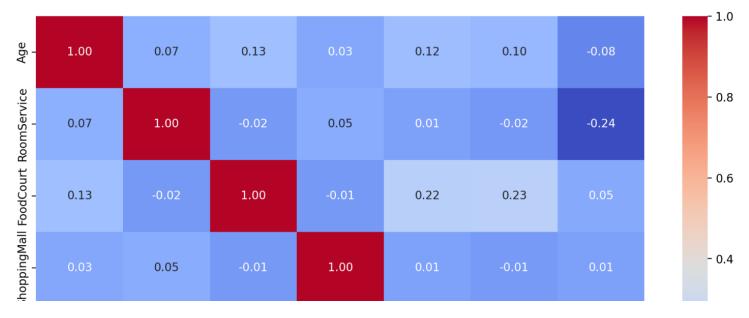
### Spa Distribution

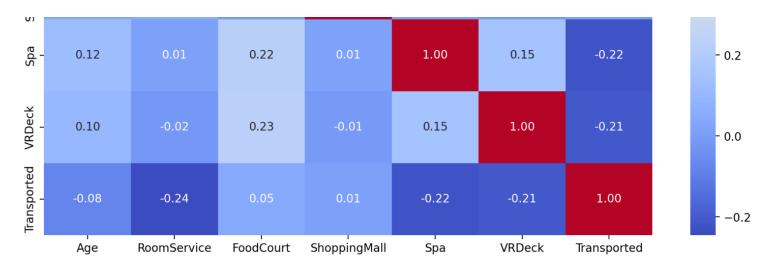


#### VRDeck Distribution



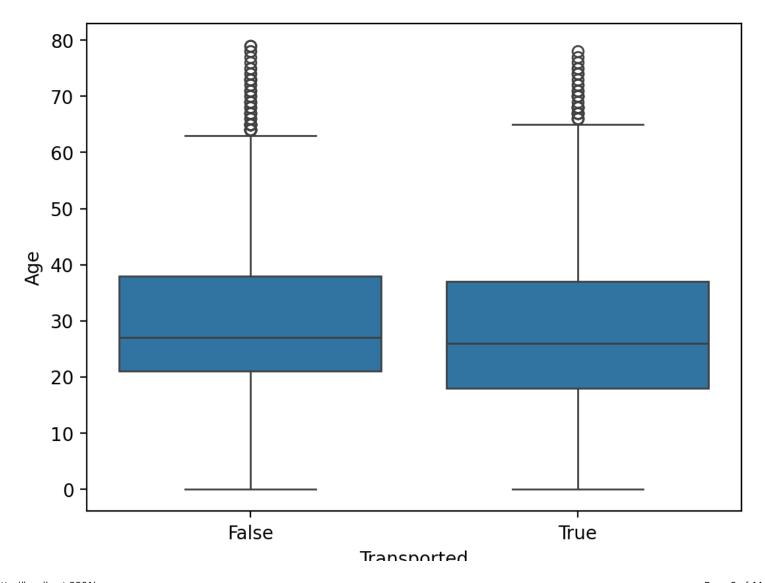
# 



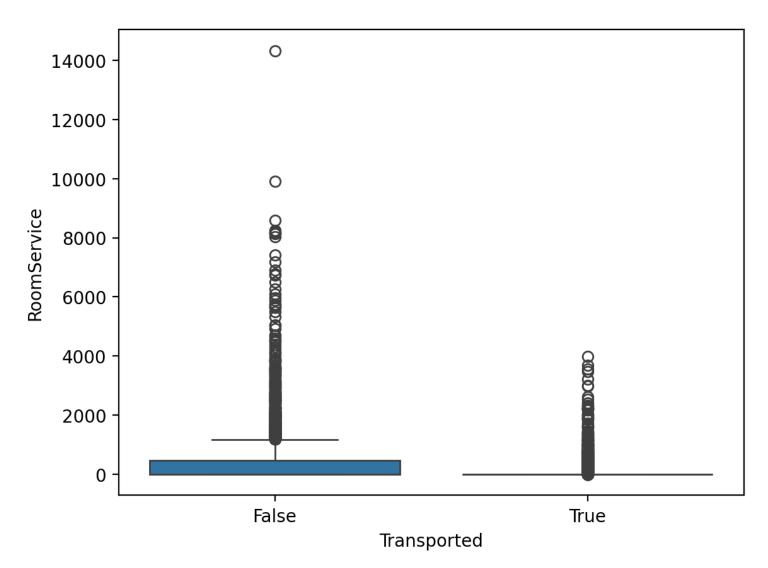


# Boxplots by 'Transported'

### **Age by Transported**

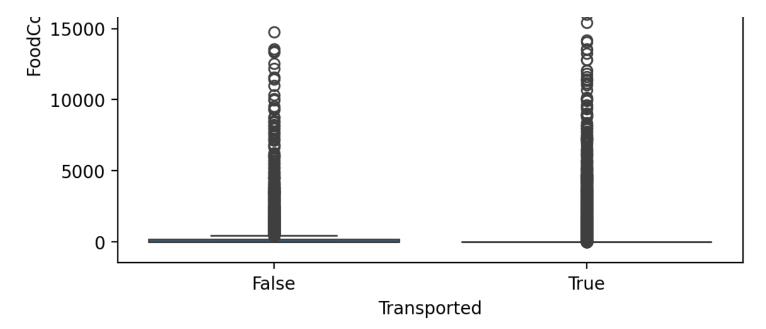


### **RoomService by Transported**

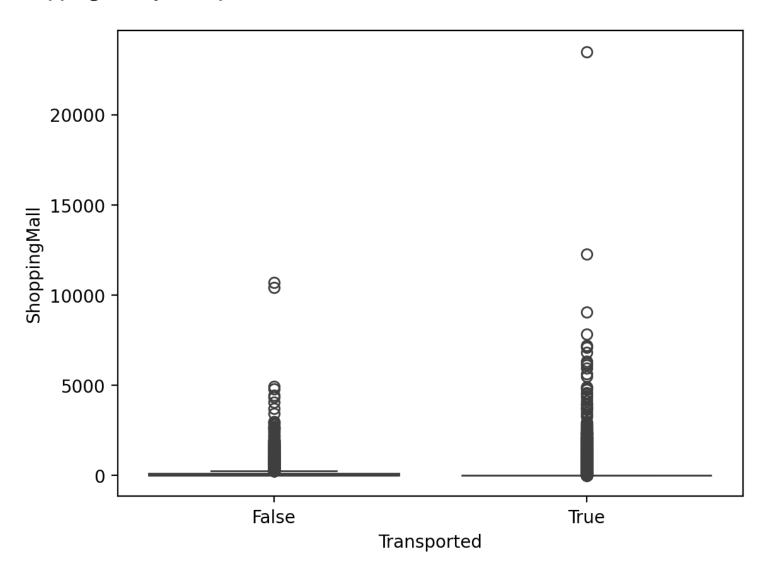


### **FoodCourt by Transported**



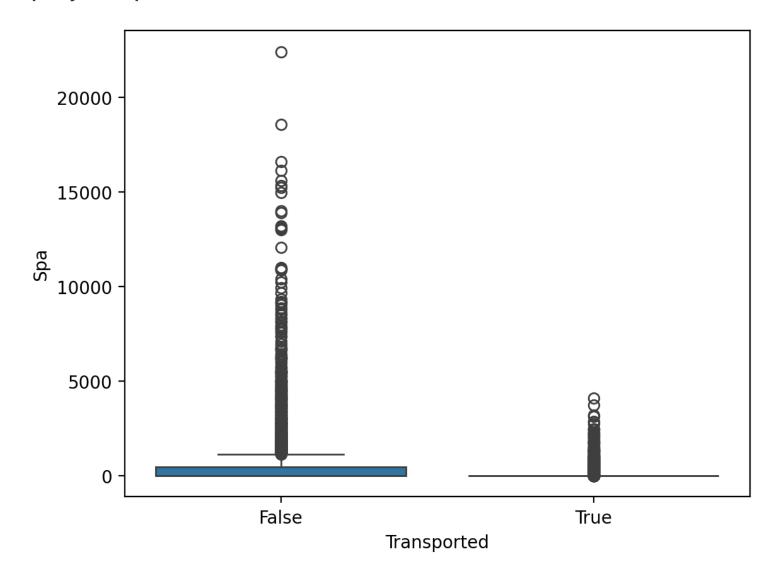


### ShoppingMall by Transported

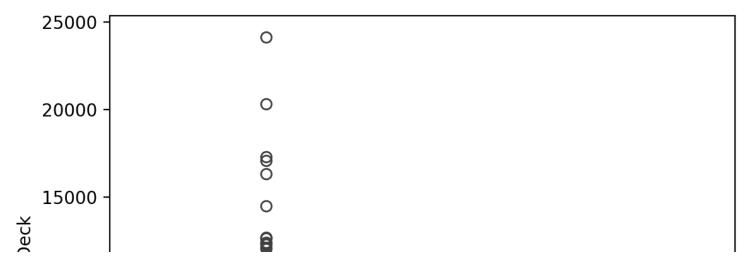


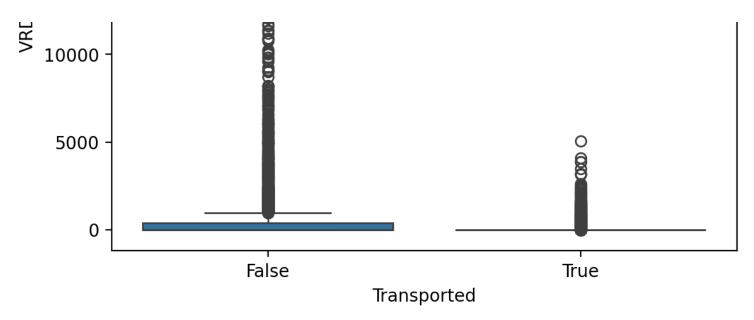
http://localhost:8501/ Page 10 of 14

### **Spa by Transported**

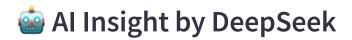


### **VRDeck by Transported**









# **Analysis of the Data Set**

The data set contains information about different types of services provided to passengers on board the Titanic. The services include RoomService, VRDeck, FoodCourt, Spa, and Transported.

# **Summary Statistics**

From the summary statistics, we can infer that:

- The average passenger had a significant number of services booked (28.83).
- Most passengers did not have any services booked, while fewer did have high volumes of services.
- There were more unique destinations compared to non-unique ones.

### **Correlations**

http://localhost:8501/ Page 12 of 14

From the top correlations we can see that:

RoomService and Transported are highly positively correlated (0.24). This implies that a passenger
who booked room service is likely to have transported as well.

 VRDeck and FoodCourt are moderately positively correlated (0.23), suggesting an increase in these services could lead to increased consumption of other services like FoodCourt or Spa.

# **Missing Values Overview**

From the missing values overview, we can infer that:

- The 'HomePlanet' data is missing for 201 passengers. This is likely due to incomplete information about the passenger's home planet.
- Data on 'CryoSleep', 'VIP' and many other services are missing for a significant number of passengers, suggesting these were not collected or do not apply to them.

# **Analysis Recommendations**

To address the missing values in the data set:

- 1. The 'HomePlanet' data can be filled with default values like 'Earth'.
- 2. For services like 'CryoSleep', it may make sense to fill these columns with a default value of 'False', indicating that the passenger did not have Cryosleep but we do not know why exactly they were missing this information.
- 3. Similar analysis can be done for other missing data points as well, based on their importance and context in the dataset.

Finally, to improve efficiency, it would make sense to categorize 'HomePlanet' into multiple binary variables (IsEarth, IsMars etc.) to simplify future analysis processes. For 'VIP', a simple binary variable can be created to help with any machine learning models.

http://localhost:8501/ Page 13 of 14

http://localhost:8501/ Page 14 of 14