2025/3/29, 08:41 CSV Data Insight

CSV Insight Generator + AI Summary

Upload a .csv file to view descriptive stats, visualizations, and AI-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 7

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

memory asage: 051.5. NE

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	Ne
freq	1	4602	5439	8	5915	None	8291	None	Ne
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	

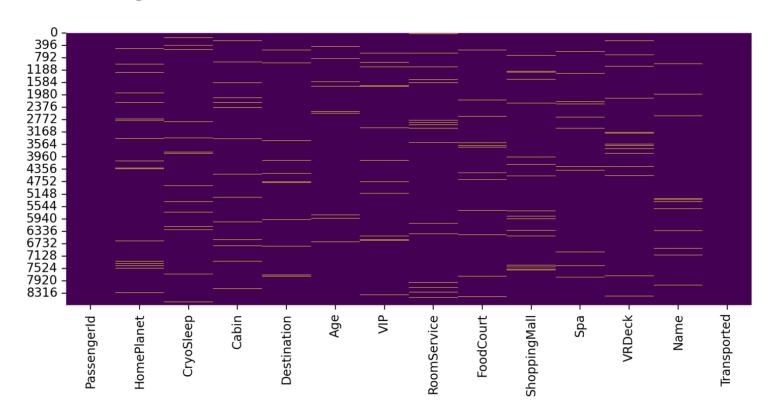
Missing Values Table

	0	
HomePlanet	20	01
CryoSleep	21	17
Cabin	19	99
Destination	18	82
Age	17	79

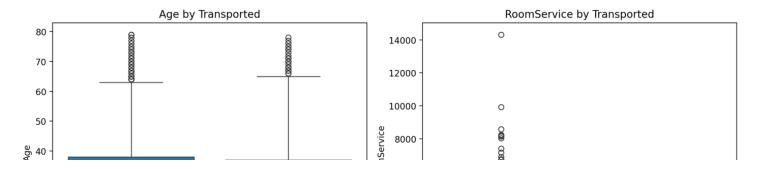
http://localhost:8501/

VIP	203
RoomService	181
FoodCourt	183
ShoppingMall	208
Spa	183

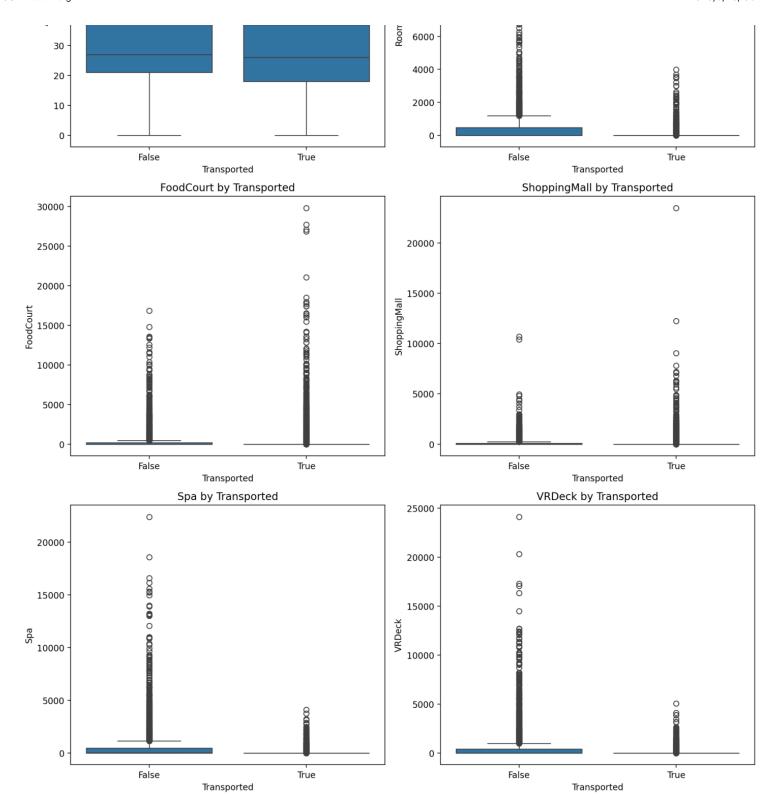
Missing Values Heatmap



Boxplots



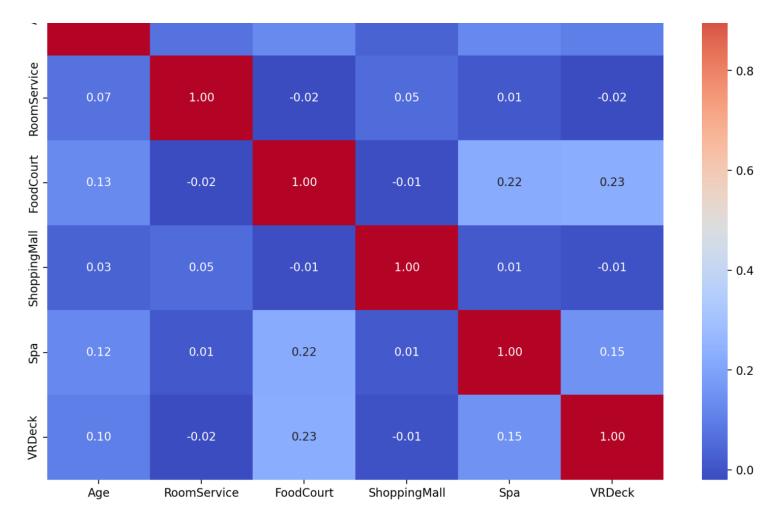
http://localhost:8501/ Page 3 of 7



Presented Services Feature Correlation Heatmap



http://localhost:8501/





Al Insight by DeepSeek

The provided CSV data seems to be related to a hypothetical space-based survival game by The Spore Company.

Here are three meaningful analyses and questions for further exploration based on the available data:

- 1. **Passenger Demographics**: Analyze the age distribution, split by HomePlanet (Earth or Europa). Could you spot any trend? Are there particular age groups that have been more likely to survive?
- Transportation and Transported Status Analysis: The data includes whether passengers were in 'CryoSleep' status, if they had services at the facilities like RoomService, FoodCourt, ShoppingMall, Spa, VRDeck etc., and their transported status.

http://localhost:8501/

- What is the average expenditure on these facilities across different destinations?
- Are there certain types of passengers who are more likely to survive if they have been 'CryoSleep' or used services at any facility?
- 3. **VIP Status and Survival**: Analyze whether VIP status impacts survival rate. If a passenger is VIP, what factors could influence their chances of survival?
 - Is there a relationship between the presence of specific facilities (like RoomService, FoodCourt, etc.) with VIP status?

These questions can help in understanding how different variables might impact the chance of survival and thus aid in making predictions or inferences about potential game mechanics.

Chat with DeepSeek Al

Ask AI further questions about this dataset:

what is the key feature of the data

Submit

Your Question:

what is the key feature of the data

Al Response:

A key feature of data in computer science refers to its attributes or characteristics that distinguish it from other types of information. These features can be categorized into several categories based on their nature and usefulness for different purposes. Some of these key features include:

1. **Quantitative**: Data that is numerical, measurable and precise such as age, salary, temperature etc. It has magnitude and order.

http://localhost:8501/ Page 6 of 7

2. **Qualitative (Categorical)**: Data that is not numerical and cannot be measured like color (blue), taste (sweet), product category (toys), etc.

- 3. **Time-related**: Timestamp when the data was collected or created. This makes it useful in analyzing trends over time.
- 4. **Location-based**: Data that includes geographical coordinates, which can be used to analyze patterns at different locations.
- 5. **Hierarchical/Structured**: Structure within the data i.e., nested sets of information (like a tree structure), in databases and file systems.
- 6. **Semi-structured**: Data that is not as easily processed but contains tags or markers to indicate certain characteristics (like JSON, XML).
- 7. **Unstructured/Non-Hierarchical**: Non-tabular, unprocessed form of data containing text, images, audio and video files etc.

Understanding these key features helps in interpreting the nature of data and choosing the right method or tool for processing it.

http://localhost:8501/ Page 7 of 7