


# Data Cards Generator

Upload a CSV file



Drag and drop file here  
Limit 200MB per file • CSV

Browse files

 fetch\_titanic\_df.csv 114.9KB

✕

## Data Table

	pclass	name	sex	age	sibsp	parch	ticket	fare
0	1	Allen, Miss. Elisabeth Walton	female	29	0	0	24160	211.3
1	1	Allison, Master. Hudson Trevor	male	0.9167	1	2	113781	151.55
2	1	Allison, Miss. Helen Loraine	female	2	1	2	113781	151.55
3	1	Allison, Mr. Hudson Joshua Creighton	male	30	1	2	113781	151.55
4	1	Allison, Mrs. Hudson J C (Bessie Waldo Daniels)	female	25	1	2	113781	151.55
5	1	Anderson, Mr. Harry	male	48	0	0	19952	21.01
6	1	Andrews, Miss. Kornelia Theodosia	female	63	1	0	13502	77.9584
7	1	Andrews, Mr. Thomas Jr	male	39	0	0	112050	
8	1	Appleton, Mrs. Edward Dale (Charlotte Lamson)	female	53	2	0	11769	51.4634
9	1	Artagaveytia, Mr. Ramon	male	71	0	0	PC 17609	49.5353

## Generate Prompt-Based Data Cards

- ☒ Include basic statistics?
- ☒ Use real column names?

Select a model:

GPT-4

▼

Rows to display in prompt:

5

110

Generate Data Cards

## Generated Prompt

You are a data scientist helping to analyze the fetch\_titanic\_df.csv dataset. This dataset has a total of 1309 rows.

Below is a small preview of the data (first 0.38% rows):

pclass,name,sex,age,sibsp,parch,ticket,fare,cabin,embarked,boat,body,home.dest,survived 1.0,"Allen, Miss. Elisabeth Walton",female,29.0,0.0,0.0,24160,211.3375,B5,S,2,"St Louis, MO",1 1.0,"Allison, Master. Hudson Trevor",male,0.9167,1.0,2.0,113781,151.55,C22 C26,S,11,"Montreal, PQ / Chesterville, ON",1 1.0,"Allison, Miss. Helen Loraine",female,2.0,1.0,2.0,113781,151.55,C22 C26,S,,,,"Montreal, PQ / Chesterville, ON",0 1.0,"Allison, Mr. Hudson Joshua Creighton",male,30.0,1.0,2.0,113781,151.55,C22 C26,S,,135.0,"Montreal, PQ / Chesterville, ON",0 1.0,"Allison, Mrs. Hudson J C (Bessie Waldo Daniels)",female,25.0,1.0,2.0,113781,151.55,C22 C26,S,,,,"Montreal, PQ / Chesterville, ON",0

*Statistical Summary:* Col 'pclass' [float64]: Miss: 0 (0%) Stats: min:1.0, 25%:2.0, med:3.0, mean:2.29, std:0.84, 75%:3.0, max:3.0 Col 'name' [object]: Miss: 0 (0%) Uniq: 1307 Top: 'Connolly, Miss. Kate':0.2%, 'Kelly, Mr. James':0.2%, 'Allen, Miss. Elisabeth Walton':0.1%, 'Ilmakangas, Miss. Ida Livija':0.1%, 'Ilieff, Mr. Ylio':0.1% Rep: Unused diverse: Bonnell, Miss. Elizabeth, Kelly, Miss. Mary, Coleff, Mr. Satio, Veal, Mr. James, Doling, Miss. Elsie Col 'sex' [object]: Miss: 0 (0%) Uniq: 2 Top: 'male':64.4%, 'female':35.6% Col 'age' [float64]: Miss: 263 (20%) Stats: min:0.17, 25%:21.0, med:28.0, mean:29.88, std:14.41, 75%:39.0, max:80.0 Col 'sibsp' [float64]: Miss: 0 (0%) Stats: min:0.0, 25%:0.0, med:0.0, mean:0.5, std:1.04, 75%:1.0, max:8.0 Col 'parch' [float64]: Miss: 0 (0%) Stats: min:0.0, 25%:0.0, med:0.0, mean:0.39, std:0.87, 75%:0.0, max:9.0 Col 'ticket' [object]: Miss: 0 (0%) Uniq: 929 Top: 'CA. 2343':0.8%, '1601':0.6%, 'CA 2144':0.6%, 'PC 17608':0.5%, '347077':0.5% Rep: Unused diverse: PC 17604, C 7077, PC 17607, PC 17603, PC 17601 Col 'fare' [float64]: Miss: 1 (0%) Stats: min:0.0, 25%:7.9, med:14.45, mean:33.3, std:51.76, 75%:31.28, max:512.33 Col 'cabin' [object]: Miss: 1014 (77%) Uniq: 186 Top: 'C23 C25 C27':2.0%, 'G6':1.7%, 'B57 B59 B63 B66':1.7%, 'F4':1.4%, 'F33':1.4% Rep: Unused diverse: F G63, F38, F E46, F G73, B36 Col 'embarked' [object]: Miss: 2 (0%) Uniq: 3 Top: 'S':69.9%, 'C':20.7%, 'Q':9.4% Col 'boat' [object]: Miss: 823 (63%) Uniq: 27 Top: '13':8.0%, 'C':7.8%, '15':7.6%, '14':6.8%, '4':6.4% Rep: Unused diverse: 11, 10, 1, 12, 16 Col 'body' [float64]: Miss: 1188 (91%) Stats: min:1.0, 25%:72.0, med:155.0, mean:160.81, std:97.7, 75%:256.0, max:328.0 Col 'home.dest' [object]: Miss: 564 (43%) Uniq: 369 Top: 'New York, NY':8.6%, 'London':1.9%, 'Montreal, PQ':1.3%, 'Paris, France':1.2%, 'Cornwall / Akron, OH':1.2% Rep: Unused diverse: London New York, NY, Paris / Montreal, PQ, London / Montreal, PQ, Cornwall / Camden, NJ, Cornwall / Hancock, MI Col 'survived' [int64]: Miss: 0 (0%) Stats: min:0.0, 25%:0.0, med:0.0, mean:0.38, std:0.49, 75%:1.0, max:1.0

**Your Task:**

- Columns Description** For each column in the dataset, return an object with the following keys:
  - Type:** The underlying data type of the column (possible values: "integer", "float", "string").
  - Domain:** An array representing the allowed values or range. For categorical columns, list the acceptable strings; for numeric columns, provide [min, max].
  - FormCanonical:** A precise canonical representation with formatting rules. For example, "Integer value without decimals representing age in years" or "Lowercase, trimmed string with fixed length".
  - Purpose:** A brief explanation of what the column represents.
  - Identifier:** A designation for the column's role. Possible values: "Measure", "Enumerated", "Binary", "Date", "Attribute", "Target".
- Data Instance** Provide a concise explanation of what a single record (row) in this dataset represents.
- Dataset Description** Write a short, high-level overview of what the dataset represents and its main purpose.
- Title** Propose a concise and clear title for this dataset.
- Dataset Usage** Describe potential use cases or scenarios for which this dataset could be applied.
- Rule Definition** Identify data quality rules with the following categories (use this exact structure):
  - FunctionalDependencies**
    - Description:* Array of objects
      - determinant:** Attributes that uniquely determine others
      - dependent:** Attributes determined by the determinant
  - ConditionalFunctionalDependencies**
    - Description:* Array of objects
      - determinant:** List of attributes
      - dependent:** List of attributes
      - condition:** The condition under which this dependency holds
  - DenialConstraints**
    - Description:* Array of objects
      - constraint:** A rule or condition that must never occur

- **OtherConstraints**
  - *Description:* Array or object for any additional constraints not covered above
- 7. **Fairness** Identify any protected or sensitive variables in the dataset (e.g., race, gender, age). For each variable, specify:
  - Potentially privileged groups
  - Potentially unprivileged groups
  - Possible biases or fairness considerations

**Output Requirements:**

- Return your answer in JSON format only.
- The JSON must have exactly seven top-level keys:
  1. "ColumnsDescription": an object where each key is a column name and each value is an object with the keys "Type", "Domain", "FormCanonical", "Purpose", and "Identifier".
  2. "DataInstance": a string.
  3. "DatasetDescription": a string.
  4. "DatasetUsage": a string.
  5. "Title": a string.
  6. "Rule\_def": an object with exactly four keys: "FunctionalDependencies", "ConditionalFunctionalDependencies", "DenialConstraints", "OtherConstraints".
  7. "Fairness": an object with two keys: "favorable\_labels" (an array of favorable values) and "protected\_attributes" (an array of objects, each with "feature" and "reference\_group", where "reference\_group" is an array specifying the privileged group).

Do not provide any additional fields or commentary.

## Model Response

```
▼ {
  ▼ "ColumnsDescription" : {
    ▼ "pclass" : {
      "Type" : "float"
      ▼ "Domain" : [
        0 : 1
        1 : 3
      ]
      "FormCanonical" :
        "Float value without decimals representing passenger class"
      "Purpose" : "The passenger class on the Titanic"
      "Identifier" : "Enumerated"
    }
    ▼ "name" : {
      "Type" : "string"
      "Domain" : "Any valid string"
      "FormCanonical" :
        "Proper case, trimmed string representing passenger's name"
      "Purpose" : "The name of the passenger"
      "Identifier" : "Attribute"
    }
    ▼ "sex" : {
      "Type" : "string"
      ▼ "Domain" : [
        0 : "male"
        1 : "female"
      ]
    }
  }
}
```

```
"FormCanonical" :  
  "Lowercase, trimmed string representing passenger's gender"  
  "Purpose" : "The gender of the passenger"  
  "Identifier" : "Enumerated"  
}  
▼ "age" : {  
  "Type" : "float"  
  ▼ "Domain" : [  
    0 : 0.17  
    1 : 80  
  ]  
  "FormCanonical" : "Float value representing age in years"  
  "Purpose" : "The age of the passenger"  
  "Identifier" : "Measure"  
}  
▼ "sibsp" : {  
  "Type" : "float"  
  ▼ "Domain" : [  
    0 : 0  
    1 : 8  
  ]  
  "FormCanonical" :  
    "Float value without decimals representing number of siblings/spouses  
    aboard"  
  "Purpose" : "The number of siblings/spouses the passenger had aboard"  
  "Identifier" : "Measure"  
}  
▼ "parch" : {  
  "Type" : "float"  
  ▼ "Domain" : [  
    0 : 0  
    1 : 9  
  ]  
  "FormCanonical" :  
    "Float value without decimals representing number of parents/children  
    aboard"  
  "Purpose" : "The number of parents/children the passenger had aboard"  
  "Identifier" : "Measure"  
}  
▼ "ticket" : {  
  "Type" : "string"  
  "Domain" : "Any valid string"  
  "FormCanonical" : "String representing ticket number"  
  "Purpose" : "The ticket number of the passenger"  
  "Identifier" : "Attribute"  
}  
▼ "fare" : {  
  "Type" : "float"  
  ▼ "Domain" : [  
    0 : 0  
    1 : 512.33  
  ]  
  "FormCanonical" : "Float value representing fare in pounds"  
  "Purpose" : "The fare the passenger paid"  
  "Identifier" : "Measure"
```

```
}
▼ "cabin" : {
  "Type" : "string"
  "Domain" : "Any valid string"
  "FormCanonical" : "String representing cabin number"
  "Purpose" : "The cabin number of the passenger"
  "Identifier" : "Attribute"
}
▼ "embarked" : {
  "Type" : "string"
  ▼ "Domain" : [
    0 : "S"
    1 : "C"
    2 : "Q"
  ]
  "FormCanonical" :
  "Single uppercase character representing port of embarkation"
  "Purpose" : "The port where the passenger embarked"
  "Identifier" : "Enumerated"
}
▼ "boat" : {
  "Type" : "string"
  "Domain" : "Any valid string"
  "FormCanonical" : "String representing lifeboat number"
  "Purpose" : "The lifeboat the passenger was on, if any"
  "Identifier" : "Attribute"
}
▼ "body" : {
  "Type" : "float"
  ▼ "Domain" : [
    0 : 1
    1 : 328
  ]
  "FormCanonical" :
  "Float value without decimals representing body number for
  identification"
  "Purpose" :
  "The body number for identification if the passenger did not survive"
  "Identifier" : "Attribute"
}
▼ "home.dest" : {
  "Type" : "string"
  "Domain" : "Any valid string"
  "FormCanonical" : "String representing home/destination"
  "Purpose" : "The home or destination of the passenger"
  "Identifier" : "Attribute"
}
▼ "survived" : {
  "Type" : "integer"
  ▼ "Domain" : [
    0 : 0
    1 : 1
  ]
  "FormCanonical" :
  "Integer value without decimals representing survival status"
```

```
    "Purpose" : "Whether the passenger survived or not"
    "Identifier" : "Target"
  }
}

"DataInstance" :
"A single record in this dataset represents a passenger on the Titanic,
including their personal details, ticket information, and survival status."

"DatasetDescription" :
"This dataset contains information about passengers on the Titanic, including
personal details, ticket information, and whether they survived the sinking of
the ship. It is often used for data analysis and machine learning tasks,
particularly for classification problems."

"Title" : "Titanic Passenger Data"

"DatasetUsage" :
"This dataset could be used for a variety of tasks, including predicting
survival on the Titanic based on various features, analyzing the factors that
contributed to survival, or studying the demographics of the passengers."

▼ "Rule_def" : {
  ▼ "FunctionalDependencies" : [
    ▼ 0 : {
      ▼ "determinant" : [
        0 : "name"
      ]
      ▼ "dependent" : [
        0 : "pclass"
        1 : "sex"
        2 : "age"
        3 : "sibsp"
        4 : "parch"
        5 : "ticket"
        6 : "fare"
        7 : "cabin"
        8 : "embarked"
        9 : "boat"
        10 : "body"
        11 : "home.dest"
        12 : "survived"
      ]
    }
  ]
  ▶ "ConditionalFunctionalDependencies" : []
  ▶ "DenialConstraints" : []
  ▶ "OtherConstraints" : []
}

▼ "Fairness" : {
  ▼ "favorable_labels" : [
    0 : 1
  ]
  ▼ "protected_attributes" : [
    ▼ 0 : {
      "feature" : "sex"
      ▼ "reference_group" : [
        0 : "male"
      ]
    }
    ▼ 1 : {
```

```
    "feature" : "age"
    ▼ "reference_group" : [
      0 : "adult"
    ]
  }
]
}
```

Study run logged to model\_history\_file.json

Title:

## Data Table

Dataset Usage:

Titanic Passenger Data

Dataset Description:

This dataset contains information about passengers on the Titanic, including personal details, ticket information, and whether they survived the sinking of the ship. It is often used for data analysis and machine learning tasks, particularly for classification problems.

	parch	ticket	fare
4	2.000000	113781	151.5
5	0.000000	19952	26.5
6	0.000000	13502	77.9
7	0.000000	112050	0.0
8	0.000000	11769	51.4
9	0.000000	PC 17609	49.5
10	0.000000	PC 17757	227.5
11	0.000000	PC 17757	227.5
12	0.000000	PC 17477	69.3
13	0.000000	19877	78.8

This dataset could be used for a variety of tasks, including predicting survival on the Titanic based on various features, analyzing the factors that contributed to survival, or studying the demographics of the passengers.

Data Instance:

A single record in this dataset represents a passenger on the Titanic, including their personal details, ticket information, and survival status.

Columns Description:

- pclass**: {'Type': 'float', 'Domain': [1.0, 3.0], 'FormCanonical': 'Float value without decimals representing passenger class', 'Purpose': 'The passenger class on the Titanic', 'Identifier': 'Enumerated'}
- name**: {'Type': 'string', 'Domain': 'Any valid string', 'FormCanonical': "Proper case, trimmed string representing passenger's name", 'Purpose': 'The name of the passenger', 'Identifier': 'Attribute'}
- sex**: {'Type': 'string', 'Domain': ['male', 'female'], 'FormCanonical': "Lowercase, trimmed string representing passenger's gender", 'Purpose': 'The gender of the passenger', 'Identifier': 'Enumerated'}
- age**: {'Type': 'float', 'Domain': [0.17, 80.0], 'FormCanonical': 'Float value representing age in years', 'Purpose': 'The age of the passenger', 'Identifier': 'Measure'}
- sibsp**: {'Type': 'float', 'Domain': [0.0, 8.0], 'FormCanonical': 'Float value without decimals representing number of siblings/spouses aboard', 'Purpose': 'The number of siblings/

Fairness:

Favorable Labels:

- 1

Protected Attributes:

- {'feature': 'sex', 'reference\_group': ['male']}
- {'feature': 'age', 'reference\_group': ['adult']}

Rule Definition:

## Functional Dependencies

- Dependency 1:** name → pclass, sex, age, sibsp, parch, ticket, fare, cabin, embarked, boat, body, home.dest, survived

## Conditional Functional Dependencies

No conditional functional dependencies identified.

## Denial Constraints

No denial constraints identified.



spouses the passenger had aboard', 'Identifier':  
'Measure'}

- **parch:** {'Type': 'float', 'Domain': [0.0, 9.0],  
'FormCanonical': 'Float value without decimals  
representing number of parents/children  
aboard', 'Purpose': 'The number of parents/  
children the passenger had aboard', 'Identifier':  
'Measure'}
- **ticket:** {'Type': 'string', 'Domain': 'Any valid  
string', 'FormCanonical': 'String representing  
ticket number', 'Purpose': 'The ticket number  
of the passenger', 'Identifier': 'Attribute'}
- **fare:** {'Type': 'float', 'Domain': [0.0, 512.33],  
'FormCanonical': 'Float value representing fare  
in pounds', 'Purpose': 'The fare the passenger  
paid', 'Identifier': 'Measure'}
- **cabin:** {'Type': 'string', 'Domain': 'Any valid  
string', 'FormCanonical': 'String representing  
cabin number', 'Purpose': 'The cabin number of  
the passenger', 'Identifier': 'Attribute'}
- **embarked:** {'Type': 'string', 'Domain': ['S', 'C',  
'Q'], 'FormCanonical': 'Single uppercase  
character representing port of embarkation',  
'Purpose': 'The port where the passenger  
embarked', 'Identifier': 'Enumerated'}
- **boat:** {'Type': 'string', 'Domain': 'Any valid  
string', 'FormCanonical': 'String representing  
lifeboat number', 'Purpose': 'The lifeboat the  
passenger was on, if any', 'Identifier':  
'Attribute'}
- **body:** {'Type': 'float', 'Domain': [1.0, 328.0],  
'FormCanonical': 'Float value without decimals  
representing body number for identification',  
'Purpose': 'The body number for identification  
if the passenger did not survive', 'Identifier':  
'Attribute'}
- **home.dest:** {'Type': 'string', 'Domain': 'Any valid  
string', 'FormCanonical': 'String representing  
home/destination', 'Purpose': 'The home or  
destination of the passenger', 'Identifier':  
'Attribute'}
- **survived:** {'Type': 'integer', 'Domain': [0, 1],  
'FormCanonical': 'Integer value without  
decimals representing survival status',  
'Purpose': 'Whether the passenger survived or  
not', 'Identifier': 'Target'}

Download JSON

## Other Constraints

No other constraints identified.

## Fairness Analysis



Select a column to plot:

pclass

Number of bins:

10

150

Plot Histogram

Select a Target column:

pclass

## Generated Fairness Prompt

You are a data scientist helping to analyze the "fetch\_titanic\_df.csv" dataset.

This dataset has a total of 1309 rows.

Below is a small preview of the data (first 5 rows):

pclass	name	sex	age	sibsp	parch	ticket	fare	cabin	embarked	boat	body	home.dest	survived
3.0	Rintamaki, Mr. Matti	male	35.0	0.0	0.0	STON/O 2. 3101273	7.1250	NaN	S	NaN	NaN	NaN	0
3.0	Nakid, Mr. Sahid	male	20.0	1.0	1.0	2653 15.7417	NaN	C	C	NaN	NaN	1	3.0
NaN	Lyntakoff, Mr. Stanko	male	NaN	0.0	0.0	349235 7.8958	NaN	S	NaN	NaN	NaN	0	3.0
NaN	Ford, Mr. Arthur	male	NaN	0.0	0.0	A/5 1478 8.0500	NaN	S	NaN	NaN	Bridgwater, Somerset, England	0	3.0

Statistical Summary:

pclass	name	sex	age	sibsp	parch	ticket	fare	cabin	embarked	boat	body	home.dest	survived	count
1309.0	1309	1309	1046.0	1309.0	1309.0	1309	1308.0	295	1307	486	121.0	745	1309.0	unique
1307	2	929	186	3	27	369								
top	Connolly, Miss. Kate	male	CA.	2343	C23	C25	C27	S	13	New York, NY				
freq	2	843	11	6	914	39	64							
mean	2.294882	29.881135	0.498854	0.385027	33.295479	160.809917	0.381971	std	0.837836	14.4135				
1.041658	0.86556	51.758668	97.696922	0.486055	min	1.0	0.1667	0.0	0.0	0.0	1.0	0.0	25%	2.0
21.0	0.0	0.0	7.8958	72.0	0.0	50%	3.0	28.0	0.0	0.0	14.4542	155.0	0.0	75%
3.0	39.0	1.0	0.0	31.275	256.0	1.0	max	3.0	80.0	8.0	9.0	512.3292	328.0	1.0

The target column(s) is/are: pclass.

Your Task:

- Fairness:** Understand the context of bias and fairness, and how it applies to this dataset.
- Protected / Sensitive Variables:** Identify columns that may contain protected attributes (e.g., race, gender, age).
- Privileged / Unprivileged Groups:** Determine which groups in the dataset might be privileged versus unprivileged, especially within any protected attributes.
- Evaluation and Mitigation:** Discuss how these fairness considerations could affect model performance and decision-making. Suggest strategies or metrics (e.g., demographic parity, equal opportunity) to detect and mitigate bias.

Output Requirements:

- Provide your answers in a structured, concise format (bullet points or paragraphs).
- Focus on fairness considerations and potential bias within the dataset.

Generate Fairness Analysis

