Data Analysis Report

Introduction

This section introduces the dataset and the purpose of this analysis.

Data Overview

Load the dataset and display basic statistics and missing values.

First few rows of the data set

	Unnamed:	0	Id	1	Name	Origin	al_name	Popularity	\
0		0	94954	Hazbin Hotel		Hazbin Hotel		277.966	
1		1	1396	Breaking	Bad	Break	ing Bad	492.705	
2		2	126308	ShÅ	gun	Š	ShÅ gun	1403.663	
3		3	94605	Ar	cane		Arcane	89.432	
4		4	37854	One P	iece	ãf¯ãf³ãf	"ãf¼ã,¹	164.547	
	First_air	_da	te Vote	_average	Vote	_count			
0	18-01-	-20	24	9.037		738			
1	20-01-	-20	80	8.906		13285			
2	27-02	-20	24	8.807		279			
3	06-11-	-20	21	8.749		3668			
4	20-10-	-19	99	8.700		4368			

Shape of the data set

(1925, 8)

Describtion of data set

Get summary statistics of numerical variables

print(df.describe())

Unnamed: 0 int64 Ιd int64 Name object Original_name object Popularity float64 First_air_date object Vote_average float64 Vote_count int64

dtype: object

Amount of missing values in the dataset

Unnamed: 0 0
Id 0
Name 0
Original_name 0
Popularity 0
First_air_date 0
Vote_average 0
Vote_count 0
dtype: int64

For futher analysis purporses all none existing Numerical data has been filled with a mean value

Dataset without NaN valeus

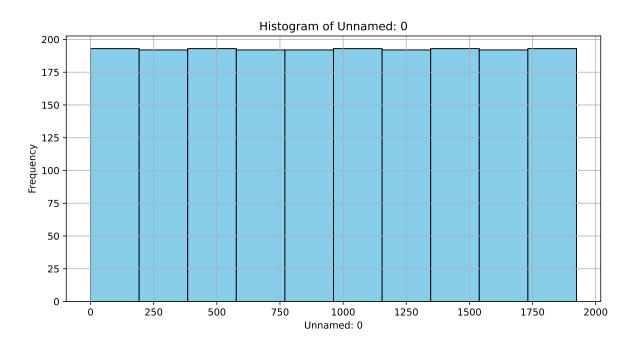
Here is how the first few rows of the cleaned dataset look:

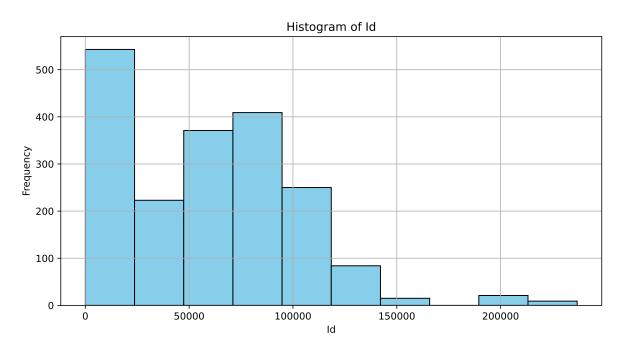
	Unnamed:	0	Id	Name	Original_name	Popularity	\
0		0	94954	Hazbin Hotel	Hazbin Hotel	277.966	
1		1	1396	Breaking Bad	Breaking Bad	492.705	
2		2	126308	ShÅ gun	ShÅ gun	1403.663	
3		3	94605	Arcane	Arcane	89.432	
4		4	37854	One Piece	ãf¯ãf³ãf"ãf¼ã.¹	164.547	

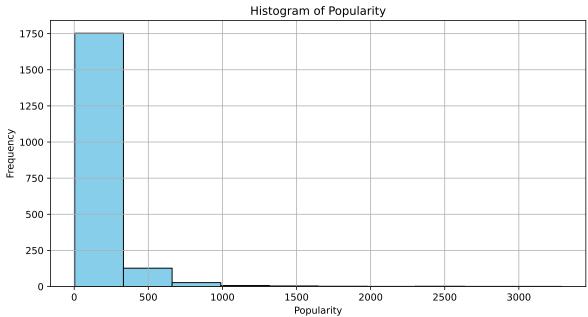
```
First_air_date
                   Vote_average
                                  Vote_count
0
      18-01-2024
                           9.037
                                          738
1
      20-01-2008
                           8.906
                                        13285
2
      27-02-2024
                           8.807
                                          279
3
      06-11-2021
                           8.749
                                         3668
4
      20-10-1999
                           8.700
                                         4368
Summary of missing values per column after cleaning:
Unnamed: 0
                   0
Ιd
                   0
Name
                   0
Original_name
                   0
                   0
Popularity
First_air_date
                   0
                   0
Vote_average
                   0
Vote_count
dtype: int64
      Unnamed: 0
                        Ιd
                                             Name
                                                            Original_name
0
                0
                    94954
                                     Hazbin Hotel
                                                             Hazbin Hotel
1
                1
                     1396
                                     Breaking Bad
                                                             Breaking Bad
2
                2
                                          ShÅ gun
                                                                  ShÅ gun
                   126308
3
                3
                                           Arcane
                    94605
                                                                   Arcane
4
                4
                    37854
                                        One Piece
                                                          ãf ãf ãf ãf ¼ã,¹
                       . . .
. . .
              . . .
1920
             1920
                     4551
                            America's Got Talent
                                                    America's Got Talent
1921
             1921
                    90755
                                       The I-Land
                                                               The I-Land
1922
             1922
                   130392
                               The D'Amelio Show
                                                       The D'Amelio Show
             1923
1923
                    10160
                                      Big Brother
                                                              Big Brother
1924
             1924
                   126725
                                            Velma
                                                                     Velma
      Popularity First_air_date
                                   Vote_average
                                                   Vote_count
0
         277.966
                       18-01-2024
                                           9.037
                                                           738
1
         492.705
                       20-01-2008
                                           8.906
                                                         13285
2
         1403.663
                       27-02-2024
                                           8.807
                                                           279
3
                       06-11-2021
                                                          3668
           89.432
                                           8.749
4
         164.547
                       20-10-1999
                                           8.700
                                                          4368
. . .
              . . .
                              . . .
                                              . . .
                                                           . . .
1920
         314.937
                       21-06-2006
                                           5.600
                                                           239
1921
           35.172
                       12-09-2019
                                           5.600
                                                           233
1922
                      03-09-2021
                                           4.987
                                                           678
           38.377
                                                           212
1923
         217.767
                       05-07-2000
                                           4.948
1924
           39.624
                       12-01-2023
                                           3.444
                                                           250
```

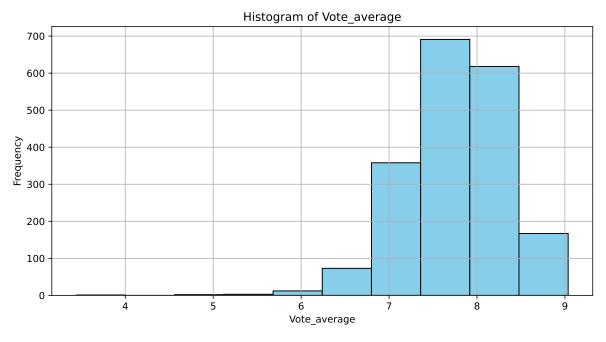
In[36]:

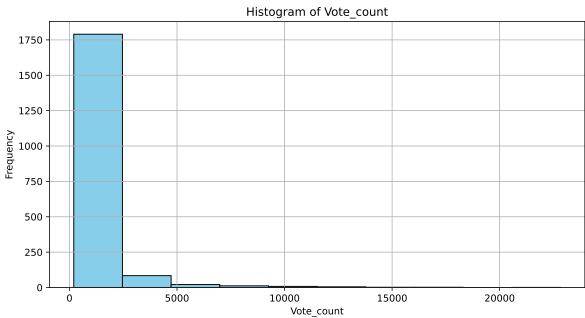
Histograms over the data











Correlation matrix over features

Correlation matrix:

Unnamed: 0 Id Popularity Vote_average Vote_count

Unnamed: 0	1.000000	-0.048941	-0.078448	-0.962084	-0.266851
Id	-0.048941	1.000000	-0.224112	0.044500	-0.070166
Popularity	-0.078448	-0.224112	1.000000	0.061174	0.297244
Vote_average	-0.962084	0.044500	0.061174	1.000000	0.252177
Vote count	-0.266851	-0.070166	0.297244	0.252177	1.000000

