

Python Summer Party Challenge

by Interview Master

Day 10 of 15

Apple

You are a Product Analyst for the Apple App Store team investigating app ratings data. Your focus is to clean and understand rating distributions across different app categories. The team wants to leverage basic statistical insights to guide app performance strategies.

Challenge Questions

Q1:

There are some data inconsistencies in the 'rating' column, specifically: leading or trailing white space, decimals represented by commas instead of decimal points (eg. 4,2 instead of 4.2), and non-numeric values. Clean up these data issues and convert the column to a numeric data type.

Q2:

Using the cleaned dataset, display the first and last five entries to get an overview of the app ratings across different categories.

Q3:

Calculate the basic summary statistics (mean, median, standard deviation) of app ratings for each category to identify variations and performance patterns.



Want to try this yourself?

Join the Challenge

Sign up for the Python Summer Party Challenge and solve 21 days
of data science problems

www.interviewmaster.ai/python-party

Or keep scrolling to see my solutions

My Solution - Q1

Day 10 Python Challenge

```
# Remove leading and trailing whitespace
cleaned_ratings = app_ratings.copy()
cleaned_ratings['rating'] = cleaned_ratings['rating'].str.strip()

# Fix decimal separators
cleaned_ratings['rating'] = cleaned_ratings['rating'].str.replace(",",".")

# Convert the cleaned strings to numeric values
cleaned_ratings['rating'] = pd.to_numeric(cleaned_ratings['rating'],
                                         errors='coerce')
print(cleaned_ratings)
```



My Solution - Q2

Day 10 Python Challenge

```
# Display the first and last five entries of the cleaned data
print(f"First five entries:\n{cleaned_ratings.head()}")
print(f"\nLast five entries:\n{cleaned_ratings.tail()}")
```



My Solution - Q3

Day 10 Python Challenge

```
# Calculate the basic summary statistics
summary_stats = cleaned_ratings.groupby('category')['rating'].agg(
    mean_rating='mean',
    median_rating='median',
    std_rating='std'
).reset_index()

print(summary_stats)
```

Ready for your own challenge?

Try this yourself by signing up for the Python Summer Party challenge:

www.interviewmaster.ai/python-party



InterviewMaster.AI