

Python Summer Party Challenge

by Interview Master

Day 6 of 15

Ben & Jerry's

You are a Product Insights Analyst working with the Ben & Jerry's sales strategy team to investigate seasonal sales patterns through comprehensive data analysis. The team wants to understand how temperature variations and unique transaction characteristics impact ice cream sales volume. Your goal is to perform detailed data cleaning and exploratory analysis to uncover meaningful insights about seasonal sales performance.

Challenge Questions

Q1:

Identify and remove any duplicate sales transactions from the dataset to ensure accurate analysis of seasonal patterns.

Q2:

Create a pivot table to summarize the total sales volume of ice cream products by month and temperature range.
Use the following temperature bins where each bin excludes the upper bound but includes the lower bound:
- Less than 60 degrees
- 60 to less than 70 degrees
- 70 to less than 80 degrees
- 80 to less than 90 degrees
- 90 to less than 100 degrees
- 100 degrees or more

Q3:

Can you detect any outliers in the monthly sales volume using the Inter Quartile Range (IQR) method? A month is considered an outlier if falls below Q1 minus 1.5 times the IQR or above Q3 plus 1.5 times the IQR.



InterviewMaster.AI

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 6 Python Challenge

```
# Note: pandas and numpy are already imported as pd and np  
# The following tables are loaded as pandas DataFrames with the  
e same names: ice_cream_sales_data  
# Please print your final result or dataframe  
ice_cream_sales_data = ice_cream_sales_data.drop_duplicates()  
print(ice_cream_sales_data)
```



InterviewMaster.AI

My Solution - Q2

Day 6 Python Challenge

```
ice_cream_sales_data = ice_cream_sales_data.drop_duplicates()

ice_cream_sales_data['sale_month'] = ice_cream_sales_data['sale_date'].dt.to_period('M').astype(str)

def temp_range(temp):
    if temp < 60:
        return '<60F'
    elif temp < 70:
        return '60-69F'
    elif temp < 80:
        return '70-79F'
    elif temp < 90:
        return '80-89F'
    elif temp < 100:
        return '90-99F'
    elif temp >= 100:
        return '100F+'
    else:
        return 'NA'

ice_cream_sales_data['temp_range'] = ice_cream_sales_data['temperature'].apply(temp_range)

pivot_df = pd.pivot_table(ice_cream_sales_data, values= 'sales_volume', index= 'sale_month', columns= 'temp_range', aggfunc= 'sum', fill_value = 0)

print(pivot_df)
```



InterviewMaster.AI

My Solution - Q3

Day 6 Python Challenge

```
ice_cream_sales_data = ice_cream_sales_data.drop_duplicates()

ice_cream_sales_data['sale_month'] = ice_cream_sales_data['sale_date'].dt.to_period('M').astype(str)

monthly_sales_df = ice_cream_sales_data.groupby('sale_month')['sales_volume'].sum().reset_index()

q1 = monthly_sales_df['sales_volume'].quantile(0.25)
q3 = monthly_sales_df['sales_volume'].quantile(0.75)
iqr = q3 - q1
lower_bound = q1 - (1.5 * iqr)
upper_bound = q3 + (1.5 * iqr)

outliers = monthly_sales_df[(monthly_sales_df['sales_volume'] < lower_bound) | (monthly_sales_df['sales_volume'] > upper_bound)]

print(outliers)
```

Ready for your own challenge?

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

www.interviewmaster.ai/python-party



InterviewMaster.AI