

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.



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

```
# Checking for duplicate data
duplicate_count = ice_cream_sales_data.duplicated().sum()
print(f"Length of data with duplicates: {len(ice_cream_sales_data)}\nDuplicate rows: {duplicate_count}")

duplicates = ice_cream_sales_data[ice_cream_sales_data.duplicated()]
print(duplicates)

# Remove duplicate data
cleaned_ice_cream_sales_data = ice_cream_sales_data.drop_duplicates()
print(cleaned_ice_cream_sales_data)
print(f"Length of data without duplicates: {len(cleaned_ice_cream_sales_data)}")
```



My Solution - Q2

Day 6 Python Challenge

```
# Extract and create the month column
cleaned_ice_cream_sales_data['sales_month'] = pd.to_datetime(cleaned_ice_cream_sales_data['sale_date']).dt.month

# Define temperature bins and labels and categorize temperatures
temp_bins = [-float('inf'), 60, 70, 80, 90, 100, float('inf')]
temp_labels = ['<60', '60-69', '70-79', '80-89', '90-99', '100+']

cleaned_ice_cream_sales_data['temp_range'] = pd.cut(
    cleaned_ice_cream_sales_data['temperature'],
    bins=temp_bins,
    labels=temp_labels,
    right=False
)

# Create the pivot table
ice_cream_pivot_table = pd.pivot_table(
    cleaned_ice_cream_sales_data,
    index='sales_month',
    columns='temp_range',
    values='sales_volume',
    aggfunc='sum',
    fill_value=0
)

print(ice_cream_pivot_table)
```



My Solution - Q3

Day 6 Python Challenge

```
# Extract and create the month column
cleaned_ice_cream_sales_data['sales_month'] = pd.to_datetime(cleaned_ice_cream_sales_data['sale_date']).dt.month

# Aggregate sales volume by month
monthly_sales_volume = (
    cleaned_ice_cream_sales_data
    .groupby('sales_month')['sales_volume']
    .sum()
)

# Calculate Q1 and Q3 and calculate the IQR
Q1 = monthly_sales_volume.quantile(0.25)
Q3 = monthly_sales_volume.quantile(0.75)
IQR = Q3 - Q1
print(f"IQR = {IQR}")

# Define outlier thresholds and identify outlier months
lower_bound = Q1 - 1.5 * IQR
upper_bound = Q3 + 1.5 * IQR
print(f"Lower bound = {lower_bound}\nUpper bound = {upper_bound}\n")

outlier_months = monthly_sales_volume[
    (monthly_sales_volume < lower_bound) | (monthly_sales_volume > upper_bound)
]

print(f"Monthly Sales Volume: \n{monthly_sales_volume}")
print(f"\nDetected Outliers: \n{outlier_months}")
```

Ready for your own challenge?

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

www.interviewmaster.ai/python-party

