Python Summer Party Challenge

by Interview Master

Day 5 of 15

Nintendo

You are a Product Analyst working with the Nintendo Switch 2 pre-sales team to analyze regional pre-order patterns and customer segmentation. Your team needs to understand how different demographics influence pre-sale volumes across regions. You will leverage historical pre-sale transaction data to extract meaningful insights that can guide marketing strategies.

Challenge Questions

Q1:

What percentage of records have missing values in at least one column? Handle the missing values, so that we have a cleaned dataset to work with.

Q2:

Using the cleaned data, calculate the total pre-sale orders per month for each region and demographic group.

Q3:

Predict the total pre-sales quantity for each region for September 2024. Assume that growth rate from August to September, is the same as the growth rate from July to August in each region.

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 5 Python Challenge

```
# Note: pandas and numpy are already imported as pd and np
# The following tables are loaded as pandas DataFrames with th
e same names: pre_sale_data
# Please print your final result or dataframe
missing_perc = (pre_sale_data.isnull().any(axis = 1).mean()) *
100
print(missing_perc)

pre_sale_data.isnull().sum()

pre_sale_data['region'] = pre_sale_data['region'].fillna('Unkn
own')
pre_sale_data['demographic_group'] = pre_sale_data['demographi
c_group'].fillna(
                                    pre_sale_data['demograph
ic_group'].mode()[0])

print(pre_sale_data)
```

InterviewMaster.AI

My Solution - Q2

Day 5 Python Challenge

```
pre_sale_data['region'] = pre_sale_data['region'].fillna('Unkn
own')
pre_sale_data['demographic_group'] = pre_sale_data['demographi
c_group'].fillna(
                                    pre_sale_data['demograph
ic_group'].mode()[0])

pre_sale_data['order_month'] = pre_sale_data['pre_order_dat
e'].dt.month

monthly_orders = pre_sale_data.groupby(['order_month', 'regio
n', 'demographic_group'])['pre_order_quantity'].sum().reset_in
dex()

print(monthly_orders)
```

InterviewMaster.AI

My Solution - Q3

Day 5 Python Challenge

```
pre_sale_data['region'] = pre_sale_data['region'].fillna('Unkn
own')
pre_sale_data['demographic_group'] = pre_sale_data['demographi
c_group'].fillna(
                                    pre_sale_data['demograph
ic_group'].mode()[0])

pre_sale_data['order_month'] = pre_sale_data['pre_order_dat
e'].dt.to_period('M').astype(str)

temp_df = pre_sale_data[pre_sale_data['order_month'].isin(['20
24-07', '2024-08'])]

grouped_df = temp_df.groupby(['region', 'order_month'])['pre_o
rder_quantity'].sum().unstack()

grouped_df = grouped_df.fillna(0)

def growth(row):
  july = row.get('2024-07', 0)
  aug = row.get('2024-08', 0)
  if july > 0:
    return (aug - july) / july
  else:
    return 0

grouped_df['growth_rate'] = grouped_df.apply(growth, axis= 1)

grouped_df['2024_September_pred'] =  grouped_df.get('2024-08',
0) * (1 + grouped_df['growth_rate'])

prediction_df = grouped_df[['2024_September_pred']].reset_inde
x()
print(prediction_df)
```