

Module 5 Challenge

[Start Assignment](#)

- Due Sep 17 by 2:59am
- Points 100
- Submitting a text entry box or a website url

Background

You'll analyze sales data to gain insights into which cities in the U.S. have sold the most athletic wear over two years. Next, you'll determine which retailers had the greatest total sales for athletic wear, and which retailers sold the most women's athletic footwear. Finally, you'll determine which day and week had the highest sales for women's athletic footwear.

Before You Begin

Before starting the assignment, be sure to complete the following steps:

- Create a new repository for this project called `athletic_sales_analysis`. **Do not add this homework assignment to an existing repository.**
- Clone the new repository to your computer.
- Push these changes to GitHub or GitLab.

Files

Download the following files to help you get started:

[Module 5 Challenge files](https://static.bc-edx.com/ai/ai-v-1-0/m5/lms/starter/M5_Starter_Code.zip)  (https://static.bc-edx.com/ai/ai-v-1-0/m5/lms/starter/M5_Starter_Code.zip)

Challenge Instructions

The starter code provided includes all the steps necessary to complete this challenge.

Combine and Clean the Data

1. Import the two CSV files, `athletic_sales_2020.csv` and `athletic_sales_2021.csv`, and read them into DataFrames.
2. Check that the columns in the two DataFrames have similar names and data types.
3. Combine the two DataFrames by the rows using an inner join, and reset the index.

SHOW HINT

4. After combining the DataFrames, do the following:
 - Check if there are any null values.
 - Check each column's data type.
 - Convert the "invoice_date" column to a datetime data type.
 - Confirm that the data type has been changed.

Determine which Region Sold the Most Products

1. Use either the `groupby` or `pivot_table` function to create a multi-index DataFrame with the "region", "state", and "city" columns.

2. Rename the aggregated column to reflect the aggregation of the data in the column.
3. Sort the results in descending order to show the top five regions, including the state and city that have the greatest number of products sold. Your final table should look like the following image:

			Total_Products_Sold
region	state	city	
Northeast	New York	New York	111954
South	Texas	Houston	90322
West	California	San Francisco	85478
		Los Angeles	76384
Southeast	Florida	Miami	73135

Determine which Region had the Most Sales

1. Use either the `groupby` or `pivot_table` function to create a multi-index DataFrame with the "region", "state", and "city" columns.
2. Rename the aggregated column to reflect the aggregation of the data in the column.
3. Sort the results in descending order to show the top five regions, including

the state and city that generated the most sales. Your final table should look like the following image:

			Total_Sales
region	state	city	
Northeast	New York	New York	39801235
West	California	San Francisco	33973228
Southeast	Florida	Miami	31600863
	South Carolina	Charleston	29285637
	Florida	Orlando	27682851

Determine which Retailer had the Most Sales

1. Use either the `groupby` or `pivot_table` function to create a multi-index DataFrame with the "retailer", "region", "state", and "city" columns.
2. Rename the aggregated column to reflect the aggregation of the data in the column.
3. Sort the results in descending order to show the top five retailers along with their region, state, and city that generated the most sales. Your final table should look like the following image:

				Total_Sales
retailer	region	state	city	
West Gear	West	California	San Francisco	32794405
Kohl's	West	California	Los Angeles	25127160
Foot Locker	Northeast	New York	New York	25008568
West Gear	West	Washington	Seattle	24862675
Foot Locker	Southeast	South Carolina	Charleston	24822280

Determine which Retailer Sold the Most Women's Athletic Footwear

1. Filter the combined DataFrame to create a DataFrame with only women's athletic footwear sales data.

SHOW HINT

2. Use either the `groupby` or `pivot_table` function to create a multi-index DataFrame with the "retailer", "region", "state", and "city" columns.
3. Rename the aggregated column to reflect the aggregation of the data in the column.

- Sort the results in descending order to show the top five retailers along with their region, state, and city that sold the most women's athletic footwear. Your final table should look like the following image:

Womens_Footwear_Units_Sold				
retailer	region	state	city	
West Gear	West	California	San Francisco	12107
Foot Locker	Northeast	New York	New York	10996
Kohl's	West	California	Los Angeles	10826
Foot Locker	Southeast	South Carolina	Charleston	8814
Sports Direct	South	Texas	Dallas	8790

Determine the Day with the Most Women's Athletic Footwear Sales

- Create a pivot table with the "invoice_date" column as the index and the "total_sales" column as the `values` parameter.
- Rename the aggregated column to reflect the aggregation of the data in the column.
- Apply the `resample` function to the pivot table, place the data into daily bins, and get the total sales for each day.
- Sort the resampled DataFrame in descending order to show the top 10 days that generated the most women's athletic footwear sales. Your final table should look like the following image:

Total Sales	
invoice_date	
2021-07-16	1521825
2021-12-16	1473497
2021-06-17	1376988
2021-08-17	1086294
2021-07-23	1021806
2021-11-17	1021145
2021-12-09	915011
2021-06-24	884238
2021-07-09	869054
2021-08-10	839120

Determine the Week with the Most Women's Athletic Footwear Sales

1. Apply `resample` to the pivot table above, place the data into weekly bins, and

get the total sales for each week.

2. Sort the resampled DataFrame in descending order to show the top 10 weeks that generated the most women's athletic footwear sales. Your final table should look like the following image:

Total Sales	
invoice_date	
2021-12-19	3098970
2021-12-12	2922161
2021-07-11	2835078
2021-07-18	2801449
2021-11-14	2531721
2021-08-22	2491259
2021-08-15	2463941
2021-11-21	2449537
2021-05-16	2422132
2021-06-13	2358602

Hints and Considerations

- Consider what you've learned so far. You've learned how to combine data using concatenation, joins, and merging, and how to reshape data using `groupby`, `pivot`, `pivot_table`, `resample`, and `melt` functions.
- If you're struggling with how to start, look back on some of the activities you did in class.
- Always commit your work and back it up with pushes to GitHub or GitLab. You don't want to lose hours of your hard work! Also make sure that your repo has a detailed `README.md` file.

Requirements

Combine and Clean the Data (15 points)

- The two DataFrames have been combined on the rows using an inner join and the index has been reset. (10 points)
- The "invoice_date" column has been converted to a datetime data type. (5 points)

Determine which Region Sold the Most Products (15 points)

- A `groupby` or `pivot_table` function has been used to create a multi-index DataFrame with the "region", "state", and "city" columns. (10 points)
- The aggregated column has been renamed to reflect the aggregation of the data in the column. (1 point)
- The results are sorted in descending order to show the top five regions, including the state and city that sold the most products. (4 points)

Determine which Region had the Most Sales (15 points)

- A `groupby` or `pivot_table` function has been used to create a multi-index DataFrame with the "region", "state", and "city" columns. (10 points)
- The aggregated column has been renamed to reflect the aggregation of the data in the column. (1 point)
- The results are sorted in descending order to show the top five regions, including the state and city that generated the most sales. (4 points)

Determine which Retailer had the Most Sales (15 points)

- A `groupby` or `pivot_table` function has been used to create a multi-index DataFrame with the "retailer", "region", "state", and "city" columns. (10 points)
- The aggregated column has been renamed to reflect the aggregation of the data in the column. (1 point)
- The results are sorted in descending order to show the top five retailers along with their region, state, and city that generated the most sales. (4 points)

Determine which Retailer Sold the Most Women's Athletic Footwear (20 points)

- A filtered DataFrame is created that shows only women's athletic footwear sales data. (8 points)
- A `groupby` or `pivot_table` function has been used to create a multi-index DataFrame with the "retailer", "region", "state", and "city" columns. (7 points)
- The aggregated column has been renamed to reflect the aggregation of the data in the column. (1 point)

- The results are sorted in descending order to show the top five retailers along with their region, state, and city that had the most women's athletic footwear sales. (4 points)

Determine the Day with the Most Women's Athletic Footwear Sales (15 points)

- A pivot table is created that has the "invoice_date" column as the index and the "total_sales" column assigned to the `values` parameter. (10 points)
- The aggregated column has been renamed to reflect the aggregation of the data in the column. (1 point)
- The `resample` function is used on the pivot table, the data is placed into daily bins, and the total sales for each day is calculated. (2 points)
- The results are sorted in descending order to show the days that generated the most women's athletic footwear sales. (2 points)

Determine the Week with the Most Women's Athletic Footwear Sales (5 points)

- The `resample` function is used on the pivot table, the data is placed into weekly bins, and the total sales for each week is calculated. (3 points)
- The results are sorted in descending order to show the weeks that generated the most women's athletic footwear sales. (2 points)

Grading

This assignment will be evaluated against the requirements and assigned a grade according to the following table:

Grade	Points
A (+/-)	90+
B (+/-)	80–89
C (+/-)	70–79
D (+/-)	60–69
F (+/-)	< 60

Submission

To submit your Challenge assignment, click Submit, and then provide the URL of your GitHub repository for grading.

NOTE


You are allowed to miss up to two Challenge assignments and still earn your certificate. If you complete all Challenge assignments, your lowest two grades will be dropped. If you wish to skip this assignment, click Next, and move on to the next module.

Comments are disabled for graded submissions in Bootcamp Spot. If you have questions about your feedback, please notify your instructional staff or your Student Success Manager. If you would like to resubmit your work for an additional review, you can use the Resubmit Assignment button to upload new links. You may resubmit up to three times for a total of four submissions.

IMPORTANT

It is your responsibility to include a note in the README section of your repo specifying code source and its location within your repo. This applies if you have worked with a peer on an assignment, used code in which you did not author or create sourced from a forum such as Stack Overflow, or you received code outside curriculum content from support staff such as an Instructor, TA, Tutor, or Learning Assistant. This will provide visibility to grading staff of your circumstance in order to avoid flagging your work as plagiarized.

If you are struggling with a challenge assignment or any aspect of the academic curriculum, please remember that there are student support services available for you:

1. Ask the class Slack channel/peer support.
2. AskBCS Learning Assistants exists in your class Slack application.
3. Office hours facilitated by your instructional staff before and after each class session.
4. **[Tutoring Guidelines](https://docs.google.com/document/d/1hTldEfWhX21B_Vz9ZentkPeziu4pPfnwiZusp=sharing)** 
https://docs.google.com/document/d/1hTldEfWhX21B_Vz9ZentkPeziu4pPfnwiZusp=sharing - schedule a tutor session in the Tutor Sessions section of Bootcampspot - Canvas
5. If the above resources are not applicable and you have a need, please reach out to a member of your instructional team, your Student Success Advisor, or submit a support ticket in the Student Support section of your BCS application.

References

Sales Product Data. Available:

<https://www.kaggle.com/datasets/knightbearr/sales-product-data> 
(<https://www.kaggle.com/datasets/knightbearr/sales-product-data>)

The sales product data above was modified by edX Boot Camps LLC, and is intended for educational purposes only.

© 2024 edX Boot Camps LLC