

Module 14 Challenge

[Start Assignment](#)

Due Sunday by 11:59pm **Points** 100 **Submitting** a text entry box or a website url

Background

Now that we've gotten a good idea of how to create our story, there is still some more work to be done to convince investors that a bike-sharing program in Des Moines is a solid business proposal. To solidify the proposal, one of the key stakeholders would like to see a bike trip analysis.

For this analysis, you'll use Pandas to change the "tripduration" column from an integer to a datetime datatype. Then, using the converted datatype, you'll create a set of visualizations to:

- Show the length of time that bikes are checked out for all riders and genders
- Show the number of bike trips for all riders and genders for each hour of each day of the week
- Show the number of bike trips for each type of user and gender for each day of the week.

Finally, you'll add these new visualizations to the two you created in this module for your final presentation and analysis to pitch to investors.

What You're Creating

This new assignment consists of two technical analysis deliverables and a written report to present your results. You will submit the following:

- Deliverable 1: Change Trip Duration to a Datetime Format
- Deliverable 2: Create Visualizations for the Trip Analysis
- Deliverable 3: Create a Story and Report for the Final Presentation

Files

Use the following link to download the Challenge starter code.

[Download the NYC CitiBike starter code.](https://2u-data-curriculum-team.s3.amazonaws.com/dataviz-online/module_14/NYC_CitiBike_Challenge_starter_code.ipynb) [_ \(https://2u-data-curriculum-team.s3.amazonaws.com/dataviz-online/module_14/NYC_CitiBike_Challenge_starter_code.ipynb\)](https://2u-data-curriculum-team.s3.amazonaws.com/dataviz-online/module_14/NYC_CitiBike_Challenge_starter_code.ipynb)

Before You Start

In Deliverable 1, you'll use Pandas to change the datatype of the "tripduration" column from an integer to a datetime datatype to get the time in hours and minutes. Using Tableau instead of Pandas to change the datatype of the "tripduration" column to a "Date and Time" will not produce the same visualizations in Deliverable 2. Trying to change the datatype by creating a calculated field may take more time for the less experienced Tableau user than the steps in Deliverable 1.

Once you change the datatype, you'll export the DataFrame as a CSV file to use for the visualizations in Deliverable 2. If you export the new CSV file into the same Tableau workbook you used in this module, the "Average Trip Duration" and "Bike Utilization" graphs will be greyed out, and you won't be able to use them because the datatype for the "tripduration" column has changed. Therefore, we recommend creating a new Tableau workbook to create the visualizations in Deliverable 2, and then you'll have to recreate the two visualizations from the module in this new Tableau workbook that you'll use for the Story in Deliverable 3. If you'd like to use the "Bike Utilization" graph in your Story for Deliverable 3, then we suggest creating a new column that contains the converted datatype of the "tripduration" column in Deliverable 1.

Deliverable 1: Change Trip Duration to a Datetime Format (20 points)

Deliverable 1 Instructions

Using Python and Pandas functions, you'll convert the "tripduration" column from an integer to a datetime datatype to get the time in hours, minutes, and seconds (00:00:00). After you convert the "tripduration" column to a datetime datatype, you'll export the DataFrame as a CSV file to use for the trip analysis in Deliverable 2.

Follow the instructions below to complete Deliverable 1.

1. Download the `NYC_CitiBike_Challenge_starter_code.ipynb` file into your bikesharing folder, and rename it `NYC_Citibike_Challenge.ipynb`.
2. Use the instructions below to add code where indicated by the numbered-step comments in the starter code file. We have provided the dependencies you will need for this challenge.
3. In Step 1, create a DataFrame from the `201908-citibike-tripdata.csv` file.
4. In Step 2, check the datatypes of each column in the DataFrame.
5. In Step 3, convert the "tripduration" column to a datetime datatype by passing the DataFrame column and the units inside the `to_datetime()` function.

If you'd like a hint on how to convert an integer datatype to a datetime datatype, that's totally okay. If not, that's great too. You can always revisit this later if you change your mind.

[SHOW HINT](#)

NOTE

You can create a new column that contains the conversion if you don't want to change the "tripduration" column to a datetime datatype.

6. In Step 4, check the datatypes of the DataFrame.
7. Confirm that the converted values in the "tripduration" column match the following image:

	tripduration	starttime	stoptime	start station id	start station name	start station latitude	start station longitude	end station id	end station name	end station latitude	end station longitude	bikeid	usertype
0	1970-01-01 00:06:33	2019-08-01 00:00:01.4680	2019-08-01 00:06:35.3780	531.0	Forsyth St & Broome St	40.718939	-73.992663	408.0	Market St & Cherry St	40.710762	-73.994004	35305	Subscriber
1	1970-01-01 00:10:27	2019-08-01 00:00:01.9290	2019-08-01 00:10:29.7840	274.0	Lafayette Ave & Fort Greene Pl	40.686919	-73.976682	3409.0	Bergen St & Smith St	40.686744	-73.990632	38822	Subscriber
2	1970-01-01 00:18:52	2019-08-01 00:00:04.0480	2019-08-01 00:18:56.1650	2000.0	Front St & Washington St	40.702551	-73.989402	3388.0	President St & Henry St	40.682800	-73.999904	18373	Subscriber
3	1970-01-01 00:29:40	2019-08-01 00:00:04.1630	2019-08-01 00:29:44.7940	479.0	9 Ave & W 45 St	40.760193	-73.991255	473.0	Rivington St & Chrystie St	40.721101	-73.991925	25002	Subscriber
4	1970-01-01 00:25:17	2019-08-01 00:00:05.4580	2019-08-01 00:25:23.4550	3312.0	1 Ave & E 94 St	40.781721	-73.945940	3312.0	1 Ave & E 94 St	40.781721	-73.945940	31198	Subscriber

If you'd like more information on the meaning of the timestamp in the "tripduration" column, read the Deep Dive below:

DEEP DIVE ▼

- In Step 5, export the DataFrame as a new CSV file without the index column. Use this new CSV file for Deliverable 2.

Deliverable 1 Requirements

You will earn a perfect score for Deliverable 1 by completing all requirements below:

- The data in the "tripduration" column is converted to a datetime datatype and has the correct time format (**15 pt**)
- The DataFrame is exported as a new file without the index column (**5 pt**)

Deliverable 2: Create Visualizations for the Trip Analysis (50 points)

Deliverable 2 Instructions

Using Tableau, create visualizations that show:

- How long bikes are checked out for all riders and genders.
- How many trips are taken by the hour for each day of the week, for all riders and genders.
- A breakdown of what days of the week a user might be more likely to check out a bike, by type of user and gender.

REWIND

For this deliverable, you've already done the following in this module:

- [Lesson 14.2.5:](#) Add the number of records (rides) as a color
- [Lesson 14.2.6:](#) Add the usertype to an axis
- [Lesson 14.2.7:](#) Change the time reference of a timedate column on an axis
- [Lesson 14.3.3:](#) Add the gender to an axis

Open Tableau and import the new CSV file that contains the conversion of the "tripduration" column to a datetime datatype in a new Tableau workbook.

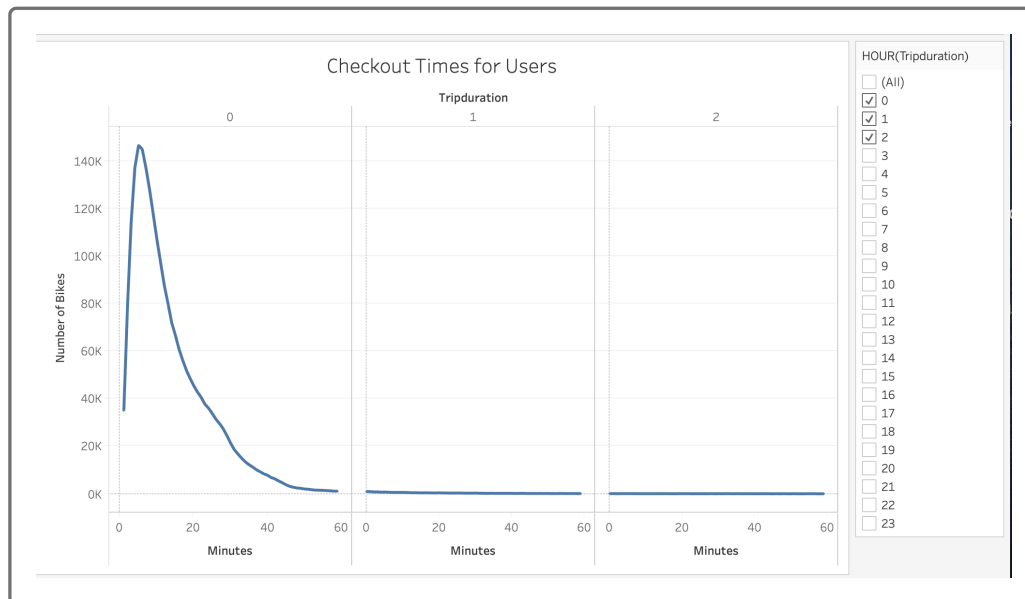
Follow the instructions below to complete Deliverable 2.

Create the Checkout Times for Users Viz

In this visualization, you'll graph the length of time that bikes are checked out for all riders.

1. Add the number of records or the generated field that counts the number of records in the CSV file to the Rows.
2. Add the "tripduration" column you converted to the Columns, and filter the "More" option by "Hour".

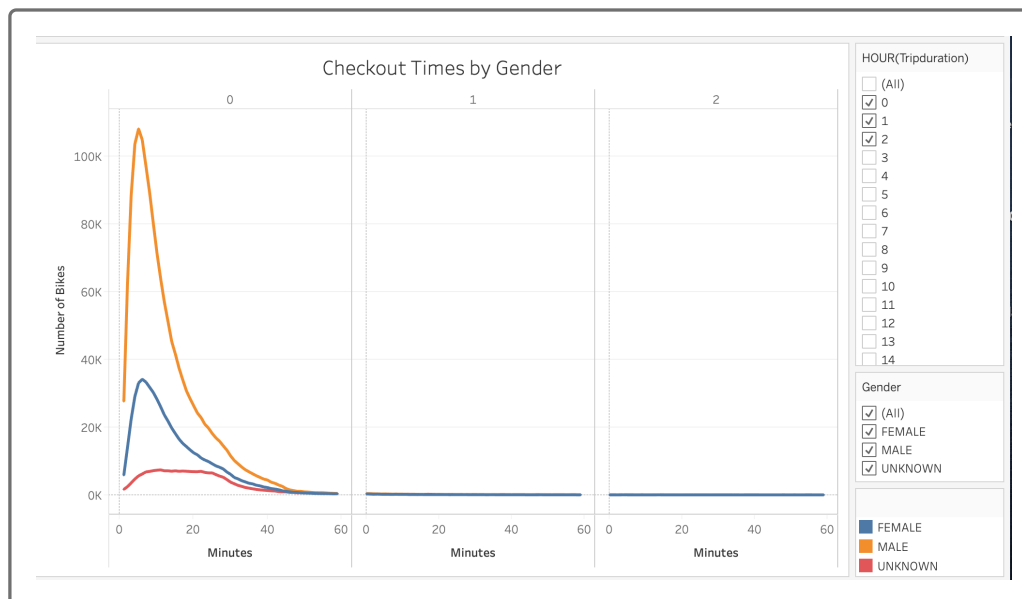
3. Add the "tripduration" column again to the Columns, and filter the "More" option by "Minute", and then change the values from "discrete" to "continuous".
4. Add the "tripduration" column that shows the "Hour" to the Filters field, and select "Show Filter".
5. Edit the X and Y axis labels by right-clicking on the axis label and selecting "Edit Axis".
6. Your graph should look similar to the following image:



Create the Checkout Times by Gender Viz

In this visualization, you'll graph the length of time that bikes are checked out for each gender.

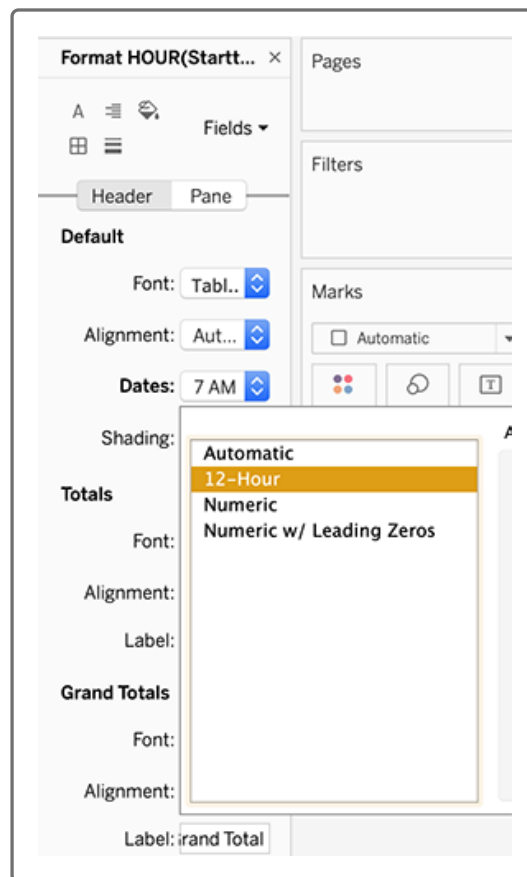
1. Repeat steps 1-4 of the "Checkout Times for Users" visualization.
2. Add the converted column for gender as a color to the Marks field, add it to the Filters field, and select "Show Filter".
3. Edit the X and Y axis labels by right-clicking on the axis label and selecting "Edit Axis".
4. Your graph should look similar to the following image:



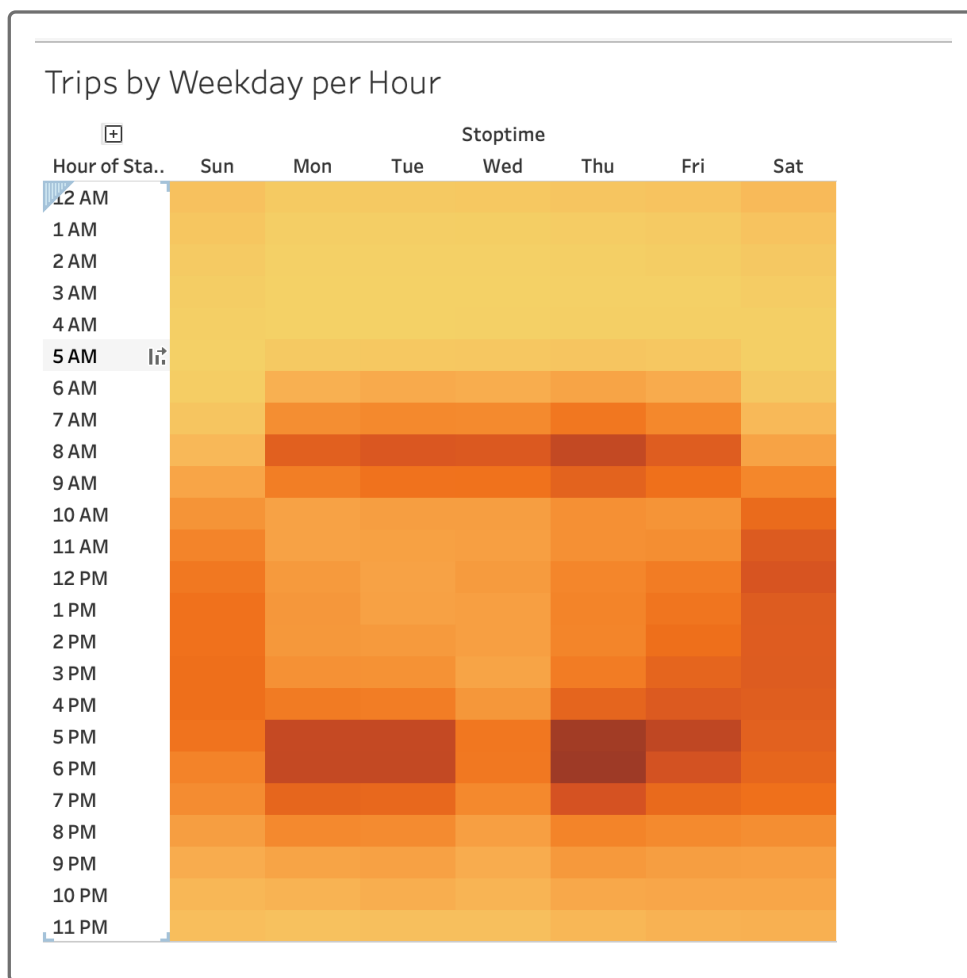
Create the Trips by Weekday for Each Hour Viz

In this visualization, you'll graph the number of bike trips by weekday for each hour of the day as a heatmap.

1. Add the "Starttime" column to the Rows, and filter the "More" option by "Hour".
2. Add the "Stoptime" column to the Columns, and filter the "More" option by "Weekday".
3. Add the number of records or the generated field that counts the number of records in the CSV file to the Marks field as a color. Select "Automatic" for the type of graph to create the heatmap.
4. Format the Y axis of the Starttime by Hour to show the 12-hour format, as shown in the following image:



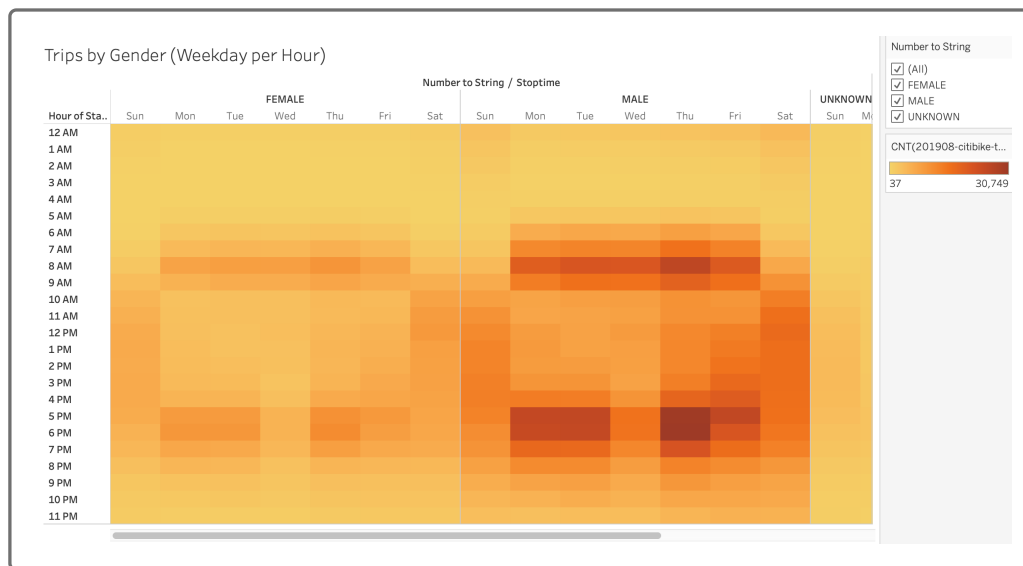
5. **Optional:** Format the X axis of Stoptime by Weekday as "Abbreviation".
6. Your graph should look similar to the following image:



Create the Trips by Gender (Weekday per Hour) Viz

In this visualization, you'll graph the number of bike trips by gender for each hour of each day of the week as a heatmap.

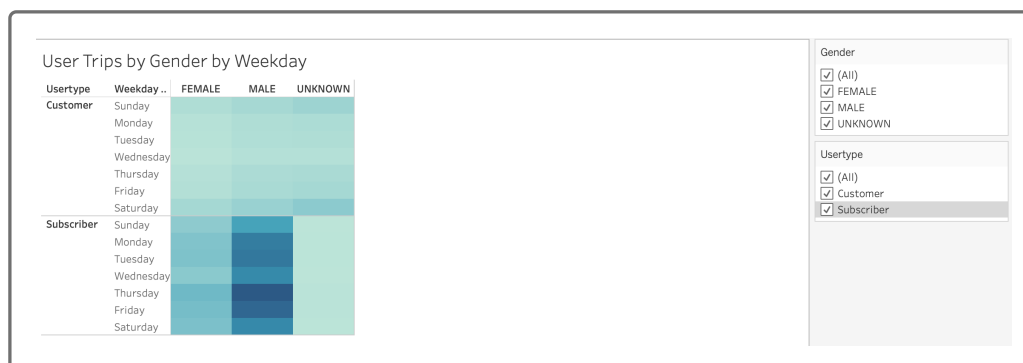
1. Repeat steps 1-3 from the "Trips by Weekday per Hour" visualization.
2. Add the converted column for "Gender" to the Columns and to the Filters field, and select "Show Filter".
3. Format the Y axis of the Starttime by Hour to show the 12-hour format.
4. **Optional:** Format the X axis of Stoptime by Weekday as "Abbreviation".
5. Your graph should look similar to the following image:



Create the User Trips by Gender by Weekday Viz

In this visualization, you'll create a heatmap that shows the number of bike trips broken down by gender for each day of the week by each Usertype.

1. Add the converted column for "Gender" to the Columns and to the Filters field, and select "Show Filter".
2. Add the "Usertype" to the Rows and to the Filters field, and select "Show Filter".
3. Add the "Starttime" column to the Rows, and filter the "More" option by "Weekday".
4. Add the number of records or the generated field that counts the number of records in the CSV file to the Marks field as a color. Select "Automatic" for the type of graph to create the heatmap.
5. Your graph should look similar to the following image:



Deliverable 2 Requirements

You will earn a perfect score for Deliverable 2 by completing all requirements below:

- There is a line graph displaying the number of bikes checked out by duration for all users, and the graph can be filtered by the hour **(10 pt)**
- There is a line graph displaying the number of bikes that are checked out by duration for each gender by the hour, and the graph can be filtered by the hour and gender **(10 pt)**
- A heatmap is created showing the number of bike trips for each hour of each day of the week **(10 pt)**
- A heatmap is created showing the number of bike trips by gender for each hour of each day of the week, and the heatmap can be filtered by gender **(10 pt)**
- A heatmap is created showing the number of bike trips for each type of user and gender for each day of the week, and you can only filter by user and gender **(10 pt)**

Deliverable 3: Create a Story and Report for the Final Presentation (30 points)

Deliverable 3 Instructions

For this part of the Challenge, you'll create a story in Tableau and write a report that describes the key outcomes of the NYC Citibike analysis you did in the module and in Deliverable 2.

REWIND

For this deliverable, you've already done the following in this module:

- [Lesson 14.5.1](#): Create a story

Follow the instructions below to complete Deliverable 2.

In Tableau, create a new Story using visualizations that will support the key findings you want to show.

1. You must use the five visualizations that you created in Deliverable 2.
2. You must use at least two visualizations that you created in this module.
3. In your README markdown file, include the following:
 - **Overview of the analysis:** Explain the purpose of this analysis.
 - **Results:** Using the visualizations you have in your Tableau Story, describe the results of each visualization underneath the image.
 - **Summary:** Provide a high-level summary of the results and two additional visualizations that you would perform with the given dataset.

Deliverable 3 Requirements

You will earn a perfect score for Deliverable 3 by completing all requirements below:

Structure, Organization, and Formatting (6 points)

The written analysis has the following structure, organization, and formatting:

- There is a title, and there are multiple sections. (2 pt)
- Each section has a heading and subheading. (2 pt)
- Links to images are working and displayed correctly. (2 pt)

Analysis (24 points)

The written analysis has the following:

1. Overview of the statistical analysis:
 - The purpose of the analysis is well defined. (5 pt)

2. Results:

- There are at least seven visualizations for the NYC Citibike analysis **(7 pt)**
- There is a description of the results for each visualization **(7 pt)**

3. Summary:

- There is a high-level summary of the results and two additional visualizations are suggested for future analysis **(5 pt)**

Submission

Once you're ready to submit, make sure to check your work against the rubric to ensure you are meeting the requirements for this Challenge one final time. It's easy to overlook items when you're in the zone!

As a reminder, the deliverables for this Challenge are as follows:

- Deliverable 1: Change Trip Duration to a Datetime Format
- Deliverable 2: Create Visualizations for the Trip Analysis
- Deliverable 3: Create a Story and Report for the Final Presentation

Upload the following to your bikesharing GitHub pages repository:

- The `NYC_Citibike_Challenge.ipynb` file.
- An updated README.md that has your written analysis with visualizations. Also, embed the Tableau Public link in your README.md file. You can do this in a few ways, but here are the two most popular ones:

`[link to dashboard](LINK GOES HERE)`
`[link to dashboard](LINK GOES HERE "link to dashboard")`

To submit your challenge assignment for grading in Bootcamp Spot, click Start Assignment, click the Website URL tab, then provide the URL of your

bikesharing GitHub repository, and then click Submit. Comments are disabled for graded submissions in BootCampSpot. If you have questions about your feedback, please notify your instructional staff or the Student Success Manager. If you would like to resubmit your work for an improved grade, you can use the **Re-Submit Assignment** button to upload new links. You may resubmit up to 3 times for a total of 4 submissions.

IMPORTANT

Once you receive feedback on your Challenge, make any suggested updates or adjustments to your work. Then, add this week's Challenge to your professional portfolio.

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.

Module-14 Rubric

Criteria	Ratings					Pts
Deliverable 1: Change Trip Duration to a Datetime Format	20 to >18.0 pts Demonstrating Proficiency ✓The "tripduration" column is converted to a datetime datatype and has the correct time format. ✓The DataFrame is exported as a new file with a .csv index.	18 to >16.0 pts Approaching Proficiency ✓The data in the "tripduration" column is converted to a datetime datatype, but the format is in seconds. ✓The DataFrame is exported as a new file without the index.	16 to >13.0 pts Developing Proficiency ✓The data in the "tripduration" column is converted to a datetime datatype, but the format is in milliseconds. ✓The DataFrame is exported as a new file without the index.	13 to >0.0 pts Emerging ✓The data in the "tripduration" column is converted to a datetime datatype, but the format is in nanoseconds. ✓The DataFrame is exported as a new file but has the index.	0 pts Incomplete	20 pts
Deliverable 2: Create Visualizations for the Trip Analysis	50 to >45.0 pts Demonstrating Proficiency ✓There is a line graph displaying the number of bikes checked out by duration for all users, and the graph can be filtered by the hour. ✓There is a line graph displaying the number of bikes that are checked out by duration for each gender by the hour, and the graph can be filtered by the hour and gender. ✓A heatmap is created showing the number of bike trips for each hour of each day of the week. ✓A heatmap is created showing the number of bike trips by gender for each hour of each day of the week, and the heatmap can be filtered by gender.	45 to >40.0 pts Approaching Proficiency ✓There is a line graph displaying the number of bikes checked out by duration for all users, and the graph can be filtered by the hour. ✓There is a line graph displaying the number of bikes that are checked out by duration for each gender by the hour, and the graph can be filtered by the hour OR gender. ✓A heatmap is created showing the number of bike trips for each hour of each day of the week. ✓A heatmap is created showing the number of bike trips by gender for each hour of each day of the week, and the heatmap can be filtered by gender.	40 to >35.0 pts Developing Proficiency ✓There is a line graph displaying the number of bikes checked out by duration for all users, but the graph can't be filtered. ✓There is a line graph displaying the number of bikes that are checked out by duration for each gender by the hour, but the graph cannot be filtered by hour OR gender. ✓A heatmap is created showing the number of bike trips for each hour of each day of the week. ✓A heatmap is created showing the number of bike trips by gender for each day of the week, and the heatmap can be filtered by gender.	35 to >30.0 pts Emerging ✓There is a line graph displaying the checkout times for each minute for all users. ✓There is a line graph displaying the number of bikes that are checked out by duration for each gender by the hour, but the graph cannot be filtered by gender. ✓A heatmap is created showing the number of bike trips for each hour of each day of the week. ✓A heatmap is created showing the number of bike trips by gender for each day of the week, but there is no filter for gender. ✓A heatmap is created showing the number of bike trips for each type of user OR gender only for each day of the week, and you can only filter by user AND gender.	0 pts Incomplete	50 pts
Deliverable 3: Create a Story and Report for the Final Presentation Structure, Organization, and Formatting	50 to >45.0 pts Demonstrating Proficiency The written analysis has ALL of the following: ✓There is a title, and there are multiple sections. ✓Each section has a heading and subheading. ✓There are images, which are formatted and displayed correctly.	45 to >40.0 pts Approaching Proficiency The written analysis has ALL of the following: ✓There is a title, and there are multiple sections. ✓Each section has a heading and subheading. ✓There are images, which are formatted and displayed correctly, with one or two minor errors.	40 to >35.0 pts Developing Proficiency The written analysis has ALL of the following: ✓There is a title, and there are multiple sections. AND ONE of the following: ✓Each section may have a heading and subheading. ✓There are images, which are formatted and displayed correctly, with one or two minor errors.	35 to >30.0 pts Emerging The written analysis has ALL of the following: ✓There is a title. ✓There may be a subheading for a section. ✓There are no headings for each section, but there are three sections.	0 pts Incomplete	6 pts

Criteria	Ratings					Pts
Deliverable 3: Create a Story and Report for the Final Presentation Analysis	24 to >21.0 pts Demonstrating Proficiency ✓The purpose is well defined. ✓There are SEVEN visualizations and a summary of each one. ✓There is a high-level summary of the results, and there are TWO suggested additional visualizations.	21 to >18.0 pts Approaching Proficiency ✓The purpose is well defined. ✓There are SIX of SEVEN visualizations and a summary of each one. ✓The high-level summary doesn't summarize all of the results, but there are TWO suggested additional visualizations.	18 to >15.0 pts Developing Proficiency ✓The purpose is well defined. ✓There are FIVE of SEVEN visualizations and a summary of each one. ✓The high-level summary doesn't summarize all of the results, and there are only ONE or TWO suggested additional visualizations.	15 to >0.0 pts Emerging ✓The purpose is well defined. ✓There are FOUR of SEVEN visualizations and a summary of each one. ✓The summary only suggests ONE or TWO additional visualizations.	0 pts Incomplete	24 pts
Total Points: 100						

© 2020 - 2022 Trilogy Education Services, a 2U, Inc. brand. All Rights Reserved.