# Generating and Interpreting Summary Statistics [Knowledge Check 1.1]

**Due** Apr 3 at 4:29pm          **Points** 7          **Questions** 7

**Available** after Mar 27 at 1pm          **Time Limit** None          **Allowed Attempts** 3

# Instructions

✅ 📄 (https://classroom.emeritus.org/courses/9054/modules/items/1506891)

✅ 🗨 (https://classroom.emeritus.org/courses/9054/modules/items/1506893)

✅ 🗨 (https://classroom.emeritus.org/courses/9054/modules/items/1506894)

✅ 🗨 (https://classroom.emeritus.org/courses/9054/modules/items/1506895)

✅ 🗨 (https://classroom.emeritus.org/courses/9054/modules/items/1506896)

📍 ✅ 🚀 (https://classroom.emeritus.org/courses/9054/modules/items/1506897)

✅ 📄 (https://classroom.emeritus.org/courses/9054/modules/items/1506899)

✅ 🗨 (https://classroom.emeritus.org/courses/9054/modules/items/1506900)

✅ 🚀 (https://classroom.emeritus.org/courses/9054/modules/items/1506901)

✅ 📄 (https://classroom.emeritus.org/courses/9054/modules/items/1506903)

✅ 📄 (https://classroom.emeritus.org/courses/9054/modules/items/1506904)

✅ 📄 (https://classroom.emeritus.org/courses/9054/modules/items/1506905)

*There is one knowledge check on this page. Please scroll down to complete it.*

# Knowledge Check 1.2: Generating and Interpreting Summary Statistics [10–15 minutes]

## 🎯 Learning Outcome Addressed:

- I~~dentify P~~ython ~~tools~~ for generating summary statistics.
- ~~~~~~~~istics.

Live Support

***This is a required activity and will count toward course completion.***

In this knowledge check, you will reflect on the Optima vs. Quanta case study and answer the following questions about generating and interpreting summary statistics.

To open this file in Google Colab, please follow these steps:

1. Click this link: **Knowledge Check 1.1** ⤷ **(https://mo-pcco.s3.us-east-1.amazonaws.com/MO-PCDS/module1/activity-name-2_starter.zip)**
2. You will be redirected to a new tab where a corresponding .zip file will download.
3. Unzip the .zip file to extract the activity's files.
4. Finally, navigate to **https://colab.research.google.com (https://colab.research.google.com/)** and click the upload tab.
5. Select the file you downloaded and want to begin. You can now view the code.
   For more information on this step, please see the **Google Colab instructions (https://classroom.emeritus.org/courses/9054/pages/google-colab?module_item_id=1506888)** from the Program Tools section.

Once in the Colab, you will be asked to download the airline cluster data (located in the colab) and upload it to the second code cell by:

- Running the first code cell
- Click the choose file button
- Upload the AirlinesCluster.csv file
- Then run the next code cell

Once you've completed this first step, execute each code block, and examine the outputs. After you have spent some time getting familiar with the code segments, you will answer some short questions regarding the concepts displayed.

## Attempt History

| | Attempt | Time | Score |
|---|---|---|---|
| **KEPT** | **Attempt 3** | 1 minute | 6.5 out of 7 |
| **LATEST** | **Attempt 3** | 1 minute | 6.5 out of 7 |
| | **Attempt 2** | 2 minutes | 5.67 out of 7 |
| | **Attempt 1** | 7 minutes | 3.83 out of 7 |

Score for this attempt: **6.5** out of 7

Submitted Mar 31 at 6:38pm

This attempt took 1 minute.

## Question 1                                           1 / 1 pts

What is the Python library that was used to read in and perform operations on the dataset?

○ PyTorch

○ Tensorflow

○ OpenCv

**Correct!**   ◉ pandas

> *That is correct! pandas is a lightweight library that provides fast and flexible data structures, making data easy to work with, whereas OpenCv is a computer vision library, and Pytorch and Tensorflow are open-source ML libraries.*

## Question 2                                           1 / 1 pts

How do you find the average value for each column in your data frame?

**Correct!**   ◉ Df.mean()

○ Df.mean

○ Df.average()

○ Df.avg()

*That is correct!  Mean() is a pandas function that can be applied to a data frame to get the mean (otherwise known as the average).*

## Question 3                                                    1 / 1 pts

How do you get all of the rows of a single column of a data frame? Assume the column you're trying to retrieve is called "BALANCE."

○ Df."BALANCE"

○ Df.BALANCE()

**Correct!**

◉ Df["BALANCE"]

○ Df[BALANCE]

*That is correct! Df["BALANCE"] is correct because you are indexing the data frame at "BALANCE."*

## Question 4                                                    0.5 / 1 pts

What part of the airline delay data can be easily overlooked and skew results?

*Select all that apply*

orrect Answer

☐ Number of flights to different airports

☐ Number of flights each day

☐ Time of day of each flight

**Correct!**          ☑ Negative numbers representing flights that arrive early

☐ Flights delayed by less than 20 minutes

---

## Question 5                                          1 / 1 pts

What did you use as a categorical variable in the flight data?

*Select all that apply*

**Correct!**          ☑ Destination airport

☐ Scheduled arrival time

☐ Actual arrival time

**Correct!**          ☑ Origin airport

☐ Arrived delayed (in minutes)

> *That is correct! The flights were categorized by route, so origin and destination airports are the categorical variables.*

---

## Question 6                                          1 / 1 pts

If you are flying from LGA to BOS and you had to choose between Optima and Quanta, you would choose Optima if you want the greatest chance of arriving on time.

**Correct!**

   ◉ True

   ○ False

> *That is correct! That answer is "true" because, when looking at the specific route, Optima has the lowest percentage of delays.*

---

## Question 7                                                    **1 / 1 pts**

Select all of the true statements.

   ☐ Optima's scheduled times are the same as Quanta's on each route.

**Correct!**

   ☑ Optima has a better performance of arriving on time on every route.

**Correct!**

   ☑ Optima has a higher percentage of flights that arrive on time.

**Correct!**

   ☑ Quanta has a lower percentage of delayed flights.

   ☐ Optima has a lower percentage of delayed flights.

*That is correct! The percentage of delayed flights is a flawed measure because it ignores cancellations and diversions. This accounts for schedule composition and differences. Initially, you found quanta that does have a lower percentage of delayed flights, although you found this metric to be deceiving.*

Quiz Score: **6.5** out of 7