



amadeus

Data Science Test

Innovation and
Research, Amadeus



Instructions

- This test is for the Data Scientist Job Position at Amadeus Innovation and Research Division
Data Scientist - Analysis and Research (R&DSB-15-01) (29841)
<https://career012.successfactors.eu/sfcareer/jobreqcareer?jobId=29841&company=AmadeusProd&username>
- Objective of the test:
 - The objective of this test is to provide the candidates the opportunity to show their competences on data science topics.
- Due Date:
 - **July 19th, 2015** (before 23:59 CET – Central European Time)
- Format of response:
 - Free: it could be a written document (e.g., pdf) and/or a presentation (e.g., ppt), video
 - Any tool/programing language can be used, it's up to you
 - DO **NOT** Attach your file(s) to the email, please provide a link to download the files (eg., dropbox)
- Send response to (both):
 - rodrigo.acunaagost@amadeus.com
 - alejandro.mottinidoliveira@amadeus.com

Instructions

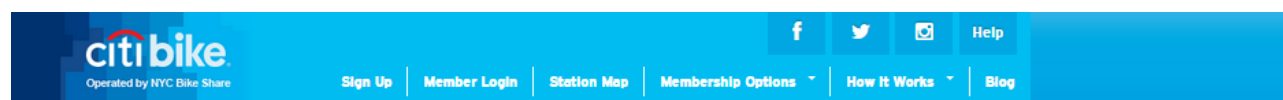
— Tasks:

- Download the datasets (Slides 4 and 5). You are free to use **one or both sets**
- Explore the data following the guideline (Slide 6)
 - Mandatory:
 - Data Collection
 - Preprocessing
 - Descriptive Analysis
 - Optional:
 - Predictive Models
 - Optimization Models
 - Advanced Visualization
- Provide a description of your process
- Do not hesitate to include any kind of chart or table to your analysis
- Do not hesitate to include any additional relevant data sources
- Answer the questions on Slide 7

Data Source 1

URL:

<http://www.citibikenyc.com/system-data>



System Data

Where do Citi Bikers ride? When do they ride? How far do they go? Which stations are most popular? What days of the week are most rides taken on? We've heard all of these questions and more from you and now we are happy to provide the data sets to help you discover the answers to these questions and more. We invite developers, engineers, statisticians, artists, academics and other members of the interested public to use the data we provide for analysis, development, visualization and whatever else moves you.

This data is provided according to the [NYCBS Data Use Policy](#).

Citi Bike Trip Histories

Below are links to downloadable files of Citi Bike trip data. The data includes:

- Trip Duration (seconds)
- Start Time and Date
- Stop Time and Date
- Start Station Name
- End Station Name
- Station ID
- Station Lat/Long
- Bike ID
- User Type (Customer = 24-hour pass or 7-day pass user; Subscriber = Annual Member)
- Gender (Zero=unknown; 1=male; 2=female)
- Year of Birth

This data has been processed to remove trips that are taken by staff as they service and inspect the system; trips that are taken to/from any of our "test" stations (which we were using more in June and July), and any trips that were below 60 seconds in length (potentially false starts or users trying to re-dock a bike to ensure it was secure).

[July 2013 \(25.8 MB\)](#)
[August 2013 \(30.6 MB\)](#)
[September 2013 \(31.6 MB\)](#)
[October 2013 \(31.5 MB\)](#)
[November 2013 \(20.6 MB\)](#)
[December 2013 \(13.6 MB\)](#)
[January 2014 \(16.2 MB\)](#)
[February 2014 \(6.9 MB\)](#)
[March 2014 \(13.4 MB\)](#)
[April 2014 \(20.4 MB\)](#)
[May 2014 \(28.3 MB\)](#)
[June 2014 \(28.5 MB\)](#)
[July 2014 \(28.4 MB\)](#)
[August 2014 \(29.2 MB\)](#)
[September 2014 \(28.8 MB\)](#)
[October 2014 \(24.9 MB\)](#)
[November 2014 \(16 MB\)](#)
[December 2014 \(12.1 MB\)](#)
[January 2015](#)
[February 2015](#)

Data Source 2

URL:

Description: http://chriswhong.com/open-data/foil_nyc_taxi/

Files: <http://www.andresmh.com/nyctaxitrips/>

NYC Taxi Trips

Data obtained through a FOIA request

[View the Project on GitHub](#)
andresmh/nyctaxitrips

View On
GitHub

Read the
story

Check the
viz

NYC Taxi Data

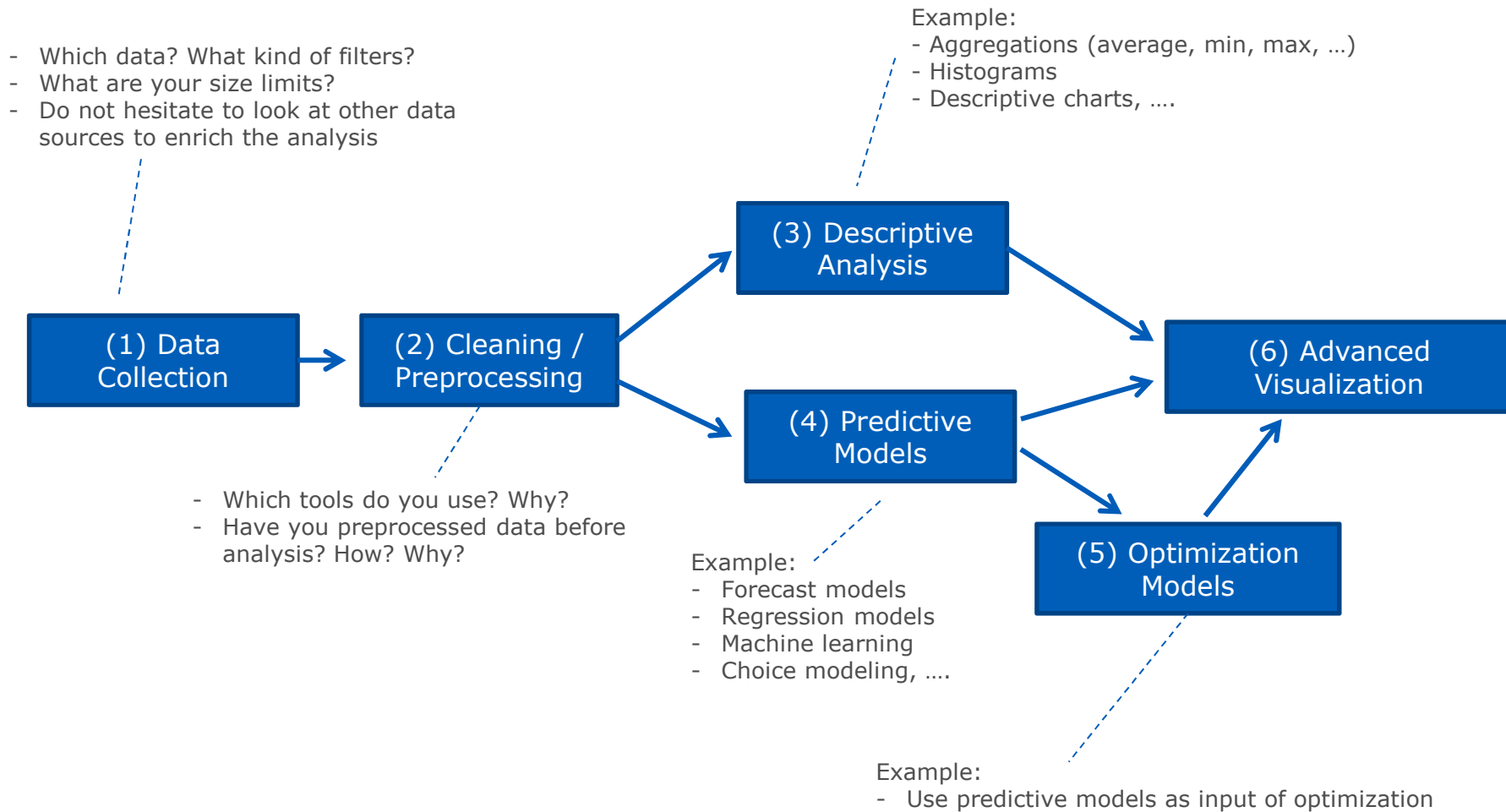
Trips

1. [trip_data_1.csv.zip](#)
2. [trip_data_2.csv.zip](#)
3. [trip_data_3.csv.zip](#)
4. [trip_data_4.csv.zip](#)
5. [trip_data_5.csv.zip](#)
6. [trip_data_6.csv.zip](#)
7. [trip_data_7.csv.zip](#)
8. [trip_data_8.csv.zip](#)
9. [trip_data_9.csv.zip](#)
10. [trip_data_10.csv.zip](#)
11. [trip_data_11.csv.zip](#)
12. [trip_data_12.csv.zip](#)

Fares

1. [trip_fare_1.csv.zip](#)
2. [trip_fare_2.csv.zip](#)
3. [trip_fare_3.csv.zip](#)
4. [trip_fare_4.csv.zip](#)
5. [trip_fare_5.csv.zip](#)
6. [trip_fare_6.csv.zip](#)
7. [trip_fare_7.csv.zip](#)
8. [trip_fare_8.csv.zip](#)
9. [trip_fare_9.csv.zip](#)
10. [trip_fare_10.csv.zip](#)
11. [trip_fare_11.csv.zip](#)
12. [trip_fare_12.csv.zip](#)

Expected Content



Some Questions

— Data Source 1:

- How many rows for October 2013?
- How many different stations for October 2013?

— Data Source 2:

- Provide Trip Time Distribution for October 2013
- What is the average total fare amount for October 2013?

— Bonus:

- What is the fastest mode of transportation, bike or taxi?