



Daulet Urazalinov <d.urazalinov@gmail.com>

mlcourse.ai. Spring 2019. Assignment #1

Google Forms <forms-receipts-noreply@google.com>
To: d.urazalinov@gmail.com

24 February 2019 at 04:39



Thanks for filling in [mlcourse.ai. Spring 2019. Assignment #1](#)

Here's what we've received from you:

EDIT RESPONSE

mlcourse.ai. Spring 2019. Assignment #1

Exploratory data analysis (EDA) of flights between US cities during year 2008 (using Pandas, Matplotlib & Seaborn).

Deadline: 2019 February 24, 20:59 GMT (London local time).

Chose one email and remember to use it in all the answer forms through the course! This will be your ID during the course. Specify your real full name in the form as well (no nicks allowed in the final top-100 rating). If in doubt, you can re-submit the form till the deadline for A1, no problem, but stick to only one email.

Email address *

Full Name *

1. How many unique carriers are there in our dataset? *

- ☐ 10
☐ 15

- ☒ 20
☐ 25

2. We have both cancelled and completed flights in the dataset. Check if there are more completed or cancelled flights. What is the difference? *

- ☐ Cancelled overweights completed by 329 flights
☒ Completed overweights cancelled by 6734860 flights
☐ Cancelled overweights completed by 671 flights
☐ Completed overweights cancelled by 11088967 flights

3. Find a flight with the longest departure delays and a flight with the longest arrival delay. Do they have the same destination airport, and if yes, what is its code? *

- ☐ yes, ATL
☐ yes, HNL
☒ yes, MSP
☐ no

4. Find the carrier that has the greatest number of cancelled flights. *

- ☐ AA
☒ MQ
☐ WN
☐ CO

5. Let's examine departure time and consider distribution by hour (column DepHour that we've created earlier). Which hour has the highest percentage of flights? *

- ☐ 1 am
☐ 5 am
☒ 8 am
☐ 3 pm

6. OK, now let's examine cancelled flight distribution by time. Which hour has the least percentage of cancelled flights? *

- ☒ 2 am
- ☐ 9 pm
- ☐ 8 am
- ☐ 3 am

7. Is there any hour that didn't have any cancelled flights at all? Check all that apply. *

- ☒ 3
- ☐ 19
- ☐ 22
- ☐ 4

8. Find the busiest hour, or in other words, the hour when the number of departed flights reaches its maximum. *

- ☐ 4
- ☐ 7
- ☒ 8
- ☐ 17

9. Since we know the departure hour, it might be interesting to examine the average delay for corresponding hour. Are there any cases, when the planes on average departed earlier than they should have done? And if yes, at what departure hours did it happen? *

- ☐ no, there are no such cases
- ☒ yes, at 5-6 am
- ☐ yes, at 9-10 am
- ☐ yes, at 2-4 pm

10. Considering only the completed flights by the carrier, that you have found in Question 4, find the distribution of these flights by hour. At what time does the greatest number of its planes depart? *

- ☐ at noon
- ☐ at 7 am
- ☐ at 8 am
- ☒ at 10 am

11. Find top-10 carriers in terms of the number of completed flights (UniqueCarrier column)? Which of the listed below is not in your top-10 list? *

- ☐ DL
- ☐ AA
- ☐ OO
- ☒ EV

12. What is the most frequent reason for flight cancellation? *

- ☐ carrier
- ☒ weather conditions
- ☐ National Air System
- ☐ security reasons

13. Which route is the most frequent, in terms of the number of flights? *

- ☐ New-York – Washington (JFK-IAD)
- ☒ San-Francisco – Los-Angeles (SFO-LAX)
- ☐ San-Jose – Dallas (SJC-DFW)
- ☐ New-York – San-Francisco (JFK-SFO)

14. Find top-5 delayed routes (count how many times they were delayed on departure). From all flights on these 5 routes, count all flights with weather conditions contributing to a delay. *

- ☐ 449
- ☐ 539
- ☐ 549
- ☒ 668

15. Choose all correct statements *

- ☐ Flights are normally distributed within time interval (0-23) (Search for: Normal distribution, bell curve).
- ☐ Flights are uniformly distributed within time interval (0-23). (Distribution is equal for all timepoints).

- ☒ In the period from 0 am to 4 am there are considerably less flights than at 7pm.

16. Show how the number of flights changes through time (on the daily/weekly/monthly basis) and interpret the findings. Choose all correct statements *

- ☒ The number of flights during weekends is less than during weekdays (working days).
- ☐ The lowest number of flights is on Sunday.
- ☒ There are less flights during winter than during summer.

17. Examine the distribution of cancellation reasons with time. Make a bar plot of cancellation reasons aggregated by months. Choose all correct statements: *

- ☒ October has the lowest number of cancellations due to weather.
- ☐ The highest number of cancellations in September is due to Security reasons.
- ☒ April's top cancellation reason is carriers.
- ☐ Flights cancellations due to National Air System are more frequent than those due to carriers.

18. Which month has the greatest number of cancellations due to Carrier? *

- ☐ May
- ☐ January
- ☐ September
- ☒ April

19. Identify the carrier with the greatest number of cancellations due to carrier in the corresponding month from the previous question. *

- ☐ 9E
- ☐ EV
- ☐ HA
- ☒ AA

20. Examine median arrival and departure delays (in time) by carrier. Which carrier has the lowest median delay time for both arrivals and departures? *

- ☐ EV

- ☐ OO
- ☐ AA
- ☒ AQ

Are there any comments on the Assignment?

Create your own Google Form