

Problem Set 3

This problem set is based on lectures 9 and 10. For a complete list of topics please consult page 2 of the course syllabus. Please consult the “Instructions for Problem Set Submissions” document under course information before submitting your assignment.

Question 1

Introduction:

Special thanks to: <https://github.com/justmarkham> for sharing the dataset and materials.

Occupations

Step 1. Import the necessary libraries

Step 2. Import the dataset from this [address](#).

Step 3. A

Step 4. D

Step 5. D

Step 6. For each occupation, calculate the

Step 7. For each combination of occupati

Step 8. For each occupation present the p

Question 2

Euro Teams

Step 1. Import the necessary libraries

Step 2. Import the dataset from this [address](#)

Step 3. Assign it to a variable called euro12

Step 4. Select only the Goal column

Step 5. How many team participated in the Euro2012?

Step 6. What is the number of columns in the dataset?

Step 7. View only the columns Team, Yellow Cards and Red Cards and assign them to a dataframe called discipline

Step 8. Sort the teams by Red Cards, then to Yellow Cards

Step 9. Calculate the mean Yellow Cards given per Team

Assignment Project Exam Help
<https://eduassistpro.github.io/>
Add WeChat: edu_assist_pro

Step 10. Filter teams that scored more than 6 goals
 Step 11. Select the teams that start with G
 Step 12. Select the first 7 columns
 Step 13. Select all columns except the last 3
 Step 14. Present only the Shooting Accuracy from England, Italy and Russia

Question 3

Housing

Step 1. Import the necessary libraries

Step 2. Create 3 different Series, each of length 100, as follows:

- The first a random number from 1 to 4
- The second a random number from 1 to 3
- The third a random number from 10,000 to 30,000

Step 3. Create a DataFrame by joining the Series by column

Step 4. Change the name of the columns to bedrs, bathrs, price_sqr_meter

Step 5. Create 3 Series and assign it to 'bigcol'

Step 6. <https://eduassistpro.github.io/>

Step 7. Reindex the DataFrame so it goes

Add WeChat edu_assist_pro

Question 4

Wind Statistics

The data have been modified to contain some missing values, identified by NaN.

Using pandas should make this exercise easier, in particular for the bonus question.

You should be able to perform all of these operations without using a for loop or other looping construct.

The data in 'wind.data' has the following format:

Yr	Mo	Dy	RPT	VAL	ROS	KIL	SHA	BIR	DUB	CLA	MUL	CLO	BEL	MAL
61	1	1	15.04	14.96	13.17	9.29	NaN	9.87	13.67	10.25	10.83	12.58	18.50	15.04
61	1	2	14.71	NaN	10.83	6.50	12.62	7.67	11.50	10.04	9.79	9.67	17.54	13.83
61	1	3	18.50	16.88	12.33	10.13	11.17	6.17	11.25	NaN	8.50	7.67	12.75	12.71

The first three columns are year, month, and day. The remaining 12 columns are average windspeeds in knots at 12 locations in Ireland on that day.

Step 1. Import the necessary libraries

Step 2. Import the dataset from this [address](#)

Step 3. Assign it to a variable called data and replace the first 3 columns by a proper datetime index.

Step 4. Year 2061? Do we really have data from this year? Create a function to fix it and apply it.

Step 5. Set the right dates as the index. Pay attention at the data type, it should be datetime64[ns].

Step 6. Compute how many values are missing for each location over the entire record. They should be ignored in all calculations below.

Step 7. Compute how many non-missing values there are in total.

Step 8. Calculate the mean windspeeds of the windspeeds over all the locations and all the times.

A single

Step 9. Compute the min, max and mean windspeeds over all the locations and all the times.

A different set of numbers for each location

Step 10. Create a DataFrame called day_windspeeds containing the min, max and mean windspeeds and standard deviations of the windspeeds across all locations for each day.

A different set of numbers for each day.

Step 11. Find the average windspeed in January for each location.

Treat January 1961 and January 1962 both as January.

Step 12. Downsample the record to a yearly frequency for each location.

Step 13. Downsample the record to a monthly frequency for each location.

Step 14. Downsample the record to a weekly frequency for each location.

Step 15. Calculate the min, max and mean windspeeds and standard deviations of the windspeeds across all locations for each week (assume that the first week starts on January 2 1961) for the first 52 weeks.

Question 5

- Step 1. Import the necessary libraries
Step 2. Import the dataset from this [address](#).
Step 3. Assign it to a variable called chipo.
Step 4. See the first 10 entries
Step 5. What is the number of observations in the dataset?
Step 6. What is the number of columns in the dataset?
Step 7. Print the name of all the columns.
Step 8. How is the dataset indexed?
Step 9. Which was the most-ordered item?
Step 10. For the most-ordered item, how many items were ordered?
Step 11. What was the most ordered item in the choice_description column?
Step 12. How many items were ordered in total?
Step 13.

- Turn the item price into a float

- Check the item price type

- C

- C

Step 14.

Step 15. How many orders were made in

Step 16. What is the average revenue am

Step 17. How many different items are so

Question 6

Create a line plot showing the number of marriages and divorces per capita in the U.S. between 1867 and 2014. Label both lines and show the legend.

Don't forget to label your axes!

Question 7

Create a vertical bar chart comparing the number of marriages and divorces per capita in the U.S. between 1900, 1950, and 2000.

Don't forget to label your axes!

Assignment Project Exam Help

<https://eduassistpro.github.io/>

Add WeChat edu_assist_pro

Question 8

Create a horizontal bar chart that compares the deadliest actors in Hollywood. Sort the actors by their kill count and label each bar with the corresponding actor's name. Don't forget to label your axes!

Question 9

Create a pie chart showing the fraction of all Roman Emperors that were assassinated.

Make sure that the pie chart is an even circle, labels the categories, and shows the percentage breakdown of the categories.

Question 10

Create a line chart showing the total revenue earned by arcades a year in the U.S. between 2000 and 2010.

Don't forget to label your axes!

Color each dot according to its year.

Assignment Project Exam Help

<https://eduassistpro.github.io/>

Add WeChat edu_assist_pro