1. Meet Your Instructors
2. Course Overview
3. Problems Solved by Data Analysts
4. Data Analysis Process Overview
5. Data Analysis Process Exercise
6. Packages Overview
7. Packages Overview Quiz
8. Tools & Environment
Downloadable resources
1. Meet Your Instructors
2. Course Overview
3. Problems Solved by Data Analysts
4. Data Analysis Process Overview
5. Data Analysis Process Exercise
6. Packages Overview
7. Packages Overview Quiz
8. Tools & Environment

Data Analysis Process Exercise

Let's build on the steps in the data analysis process with some questions you might have at each step. We'll use data from **Kaggle's Bike Sharing Demand** competition. In this dataset, you are given hourly rental data spanning two years from the **Capital Bikeshare** program in Washington, D.C. Below is a screenshot of the first five rows of this dataset.

Note: This dataset has been slightly modified from the original version.

	datetime	season	holiday	workingday	weather	temp	atemp	humidity	windspeed	casual	registered	count
0	2011-01-01 00:00:00	1	0.0	0	1	9.84	14.395	81	0.0	3	13	16
1	2011-01-01 01:00:00	1	0.0	0	1	9.02	13.635	80	0.0	8	32	40
2	2011-01-01 02:00:00	1	0.0	0	1	9.02	13.635	80	0.0	5	27	32
3	2011-01-01 03:00:00	1	0.0	0	1	9.84	14.395	75	0.0	3	10	13
4	2011-01-01 04:00:00	1	0.0	0	1	9.84	14.395	75	0.0	0	1	1

Use the table below to help you answer the first question. Each feature is a column in the dataset.

Feature	Description
datetime	hourly date + timestamp
season	1 = spring, 2 = summer, 3 = fall, 4 = winter
holiday	whether the day is considered a holiday
workingday	whether the day is neither a weekend nor holiday
weather *	1, 2, 3, 4 (see descriptions below)
temp	temperature in Celsius
atemp	"feels like" temperature in Celsius
humidity	relative humidity
windspeed	wind speed
casual	number of non-registered user rentals initiated
registered	number of registered user rentals initiated
count	number of total rentals

*** Keys for Weather Feature**

1. clear, few clouds, partly cloudy, partly cloudy
2. mist + cloudy, mist + broken clouds, mist + few clouds, mist
3. light snow, light rain + thunderstorm + scattered clouds, light rain + scattered clouds
4. heavy rain + ice pallets + thunderstorm + mist, snow + fog

Quiz Question

Question Step

Given the above data on variables that potentially influence the number of bikes rented each hour, what questions would be relevant to ask? (You may select more than one.)

- [x] Which attributes are most important in predicting the number of bikes rented?
- [ ] What type of bike is rented most frequently?
- [x] For which day of the week should the bikesharing company run promotions if the goal is to smooth out the number of rentals across the week?
- [ ] Should a given station stock more bikes in order to maximize profits?

Submit

	season	holiday	workingday	weather	temp	atemp	humidity	windspeed	casual	registered	count
count	10886.000000	9253.000000	10886.000000	10886.000000	10886.000000	10057.000000	10886.000000	10886.000000	10886.000000	10886.000000	10886.000000
mean	2.506614	0.031017	0.680875	1.418427	20.251393	23.737533	61.886460	12.799395	36.021955	155.552177	191.574132
std	1.116174	0.173373	0.466159	0.633839	8.058472	8.714964	19.245033	8.164537	49.960477	151.039033	181.144454

std	1.116174	0.178375	0.466159	0.633839	8.638472	8.714904	19.245033	8.164537	49.960477	151.039033	181.144454
min	1.000000	0.000000	0.000000	1.000000	0.820000	0.760000	0.000000	0.000000	0.000000	0.000000	1.000000
25%	2.000000	0.000000	0.000000	1.000000	13.940000	16.665000	47.000000	7.001500	4.000000	36.000000	42.000000
50%	3.000000	0.000000	1.000000	1.000000	20.500000	24.240000	62.000000	12.998000	17.000000	118.000000	145.000000
75%	4.000000	0.000000	1.000000	2.000000	26.240000	31.060000	77.000000	16.997900	49.000000	222.000000	284.000000
max	4.000000	1.000000	1.000000	4.000000	235.000000	45.455000	100.000000	56.996900	367.000000	886.000000	977.000000

Helpful stats on each column for the next question

```
RangeIndex: 10886 entries, 0 to 10885
Data columns (total 12 columns):
datetime      10886 non-null object
season        10886 non-null int64
holiday       9253 non-null float64
workingday    10886 non-null int64
weather       10886 non-null int64
temp          10886 non-null float64
atemp         10057 non-null float64
humidity      10886 non-null int64
windspeed     10886 non-null float64
casual        10886 non-null int64
registered    10886 non-null int64
count         10886 non-null int64
dtypes: float64(4), int64(7), object(1)
```

 # Non-null values and datatype for each column

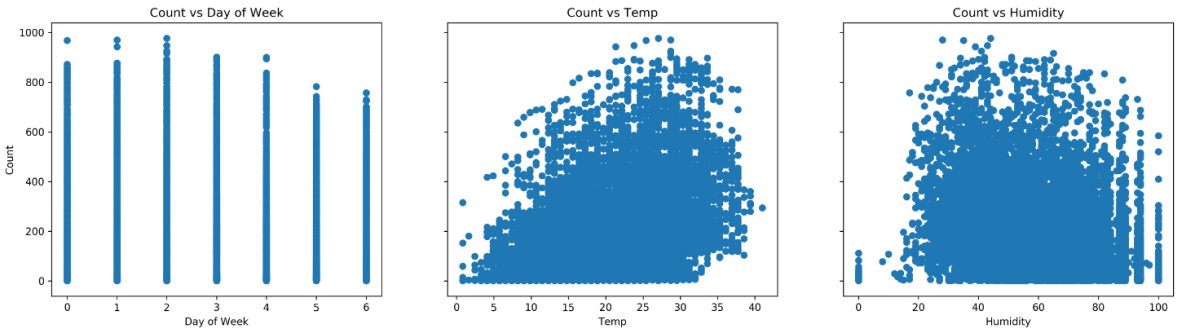
Quiz Question

Wrangle Step

What potential problems do you see with this Kaggle bike sharing dataset that would need to be fixed before continuing with analysis? (You may select more than one.)

- [x] Dates are not in date format
- [x] Some values are missing
- [x] Temperature values are far outside a realistic range on Earth
- [ ] Weather can't be represented by numerical values

Submit



Scatterplots on the number of bikes rented vs. day of week, temperature, and humidity

Quiz Question

Explore Step

Based on these scatterplots, which of these three features seems most helpful in predicting count?

- ( ) Day of Week
- (•) Temperature

Humidity

Submit

Quiz Question

Communicate Step

What would be valid methods of communicating your conclusions from the Bike Sharing data?

- [ ] Scatterplots for correlation among features such as temperature vs humidity
- [x] A written report detailing the most important variables to consider when predicting the number of bike rentals
- [ ] A graph of the regression equation for different temperatures

Submit

Great job! Next, let's learn about Python packages, which were used to prepare this exercise!

Previous Next

Give Page Feedback