



## Microsoft: DAT210x Programming with Python for Data Science



Bookmarks

- ▶ Start Here
- ▶ 1. The Big Picture
- ▶ 2. Data And Features
- ▶ 3. Exploring Data
- ▶ 4. Transforming Data
- ▼ 5. Data Modeling

Lecture: Clustering  
Quiz



Lab: Clustering  
Lab



Lecture: Splitting Data  
Quiz



Lecture: K-Nearest  
Neighbors  
Quiz



Lab: K-Nearest Neighbors

## 5. Data Modeling &gt; Lab: Clustering &gt; Assignment 3

Bookmark

## Lab Assignment 3

Continuing on with the previous lab, this time you'll validate your results by comparing the user's weekday activity to their weekend activity. To get started, use the starter code in `/Module5/assignment3.py`.

1. Load up the same CDR dataset into a dataframe, and extract the unique "In" phone numbers. You don't have to save it as a Python list this time, and can keep it as an NDArray. The previous lab had you convert to a list just so you'd have the experience doing it.
2. Create a new slice, once again for the first unique number in the CDR. Instead of limiting it to Weekend only entries, index it so that the slice only contains Weekday entries, **Mon-Fri**, and so that it occurs any time before 5pm.
3. Run K-Means on the data with K=4. Plot the cellphone towers the user connected to, and then plot the cluster centers using a different marker and color.
4. Answer the questions below.

## Lab Questions

(3/3 points)

Answer the following questions given the data you just recorded, for K=4, and CallTime is less than 5pm (that is "17:00:00"), and the call's day-of-week being a weekday.

Lab

**Lecture: Regression**

Quiz

**Lab: Regression**

Lab



Dive Deeper

- ▶ 6. Data Modeling II
- ▶ 7. Evaluating Data
- ▶ Course Wrap-up

The user's home location will likely be near the centroid with the second most attached samples. Does your approximated home location from this map coincide with the home approximation from the previous lab?

☐ Yes, they are exactly the same

☒ Yes, they match, but there is a slight difference ✓

☐ No, you can tell that they should match; however, their locations are *\*very\** different

☐ No, they are completely different

#### EXPLANATION

...

Given the indexed time range, and the times people usually receive / make calls, the cluster with the most samples is likely to be the user's work location. What is the phone number of the user who works at the US Post Office near Cockrell Hill Rd?

289-436-5987 ▼



Answer: 289-436-5987

Run your assignment with  $K=3$ . Look at the code that gets the mean CallTime value for the cluster with the least amount of samples assigned to it (the cluster we suspect corresponds to the user transiting to work). What hour is the average CallTime value of that cluster closest to?

8 am ▾



Answer: 8 am

*You have used 2 of 2 submissions*

© All Rights Reserved



© edX Inc. All rights reserved except where noted. EdX, Open edX and the edX and Open EdX logos are registered trademarks or trademarks of edX Inc.

