Jennifer Brown

May 13, 2017

Datascope Coding Challenge

**Warm Up questions**

1. Which stop has the highest average ridership per day, and what is it?

* Not accounting for holidays, the station with the highest average ridership per day is Clark & Lake on Wednesdays with 17,835 average riders (17834.659446).
* Code used to calculate this figure can be found at:
  + <https://github.com/jabrown5/CTA-Ridership/blob/master/CTA_analyses.sas>
  + See portion commented as /\* CODE TO ASSESS QUESTION 1 \*/
* A table with the stations and the highest ridership for 20 station and day combinations can be found just below:

|  |  |  |  |
| --- | --- | --- | --- |
| **Obs** | **stationname** | **dayOfWeek** | **rides** |
| **1** | Clark/Lake | Wednesday | 17834.659446 |
| **2** | Clark/Lake | Tuesday | 17802.318893 |
| **3** | Clark/Lake | Thursday | 17602.978313 |
| **4** | Clark/Lake | Friday | 17049.066265 |
| **5** | Clark/Lake | Monday | 16486.11071 |
| **6** | Lake/State | Wednesday | 15803.713598 |
| **7** | Lake/State | Friday | 15799.720482 |
| **8** | Lake/State | Tuesday | 15677.605295 |
| **9** | Lake/State | Thursday | 15638.985542 |
| **10** | Chicago/State | Friday | 15141.627711 |
| **11** | Lake/State | Monday | 14605.851986 |
| **12** | Chicago/State | Thursday | 14019.415663 |
| **13** | Chicago/State | Tuesday | 13978.661853 |
| **14** | Chicago/State | Wednesday | 13892.548736 |
| **15** | Chicago/State | Monday | 12994.154031 |
| **16** | 95th/Dan Ryan | Wednesday | 12836.375604 |
| **17** | 95th/Dan Ryan | Tuesday | 12827.045949 |
| **18** | 95th/Dan Ryan | Friday | 12656.306295 |
| **19** | 95th/Dan Ryan | Thursday | 12606.348247 |
| **20** | Chicago/State | Saturday | 12134.492771 |

1. Which stop has the greatest standard deviation in *weekday (exclude holidays)* ridership per day, and what is it?

* Excluding weekends and holidays, the station with the greatest standard deviation in ridership per day is Lake & State on Friday with a standard deviation of 4,642 riders (4642.1890859)
* Code used to calculate this figure can be found at:
  + <https://github.com/jabrown5/CTA-Ridership/blob/master/CTA_analyses.sas>
  + See portion commented as /\* CODE TO ASSESS QUESTION 2 \*/
* A table with the stations and the greatest standard deviation in ridership for 20 stations (and days) can be found just below:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Obs** | **Stationname** | **dayOfWeek** | **rides** | **std\_deviation** |
| **1** | Lake/State | Friday | 15869.389769 | 4642.1890859 |
| **2** | Lake/State | Thursday | 15963.387097 | 4272.2714117 |
| **3** | Lake/State | Wednesday | 15889.985437 | 4190.0746131 |
| **4** | Lake/State | Monday | 15068.546379 | 4151.6859602 |
| **5** | Lake/State | Tuesday | 15780.955097 | 4132.2580174 |
| **6** | Addison-North Main | Friday | 8498.5213155 | 4053.7547989 |
| **7** | Washington/State | Wednesday | 6037.361244 | 3935.494118 |
| **8** | Washington/State | Tuesday | 6038.1726619 | 3919.5840131 |
| **9** | Washington/State | Thursday | 6032.0317848 | 3894.7784183 |
| **10** | Washington/State | Friday | 5957.702381 | 3881.5531804 |
| **11** | Washington/State | Monday | 5807.5213033 | 3832.2888642 |
| **12** | Addison-North Main | Wednesday | 7550.3377724 | 3164.7133576 |
| **13** | Clark/Lake | Friday | 17179.705238 | 3094.1141119 |
| **14** | Addison-North Main | Tuesday | 7472.4563107 | 3092.4972851 |
| **15** | Addison-North Main | Thursday | 7474.5905707 | 3052.9885513 |
| **16** | Clark/Lake | Monday | 17181.796696 | 2979.8659984 |
| **17** | Fullerton | Thursday | 12161.847395 | 2729.2368988 |
| **18** | Fullerton | Wednesday | 12006.512136 | 2703.584732 |
| **19** | Addison-North Main | Monday | 6654.8754765 | 2640.3706593 |
| **20** | Roosevelt | Friday | 9452.5961071 | 2636.7045209 |

**Challenge question**

* + What questions could you potentially explore/answer with this data?

For the purposes of this challenge, I wanted to focus on the Main St station “L” stop in Evanston. While not in Chicago, the city of Evanston is serviced by Chicago’s “L” system and I live in the area and, selfishly, would like a business of my choosing to move in for my convenience. Specifically, would a Dunkin Donuts or similarly themed breakfast and coffee fast-food type establishment be a smart business move?

With the specific data indicated, a good starting point as been provided for research into the feasibility of a Dunkin Donuts (or similar) establishment.

1. We can explore overall foot-traffic to this “L” station (as measured by the station entry counts). This can give some indication as to the potential pool of customers that may pass by.

**Evanston Main Street “L” station entries:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Obs** | **stationname** | **dayOfWeek** | **rides** | **std\_deviation** |
| **6** | Main | Monday | 1138.3341804 | 101.39454537 |
| **9** | Main | Tuesday | 1206.4824242 | 88.445529252 |
| **12** | Main | Wednesday | 1186.3066667 | 92.267363018 |
| **15** | Main | Thursday | 1205.0086849 | 93.957086406 |
| **18** | Main | Friday | 1224.2691839 | 138.18684554 |

Weekday average rides since 2001 for the Evanston Main Street “L” station range from 1,138 on Mondays to 1,224 on Fridays. Fridays exhibit the greatest standard deviation, however.

1. We can determine the busiest days of the week for “L” station usage as measured by average station entries by day of the week. This could inform business hours and staffing.

The busiest day of the week for the Main Street “L” station is Friday followed closely by Tuesday and Thursday. Fridays, however, exhibit the greatest standard deviations in ridership. This is illustrated by the table above.

1. We can look for comparable station stops in the “L” system with similar average daily ridership (and standard deviation). After finding similar stops based on ridership, we can look at the neighborhood where the stop is located for similarities to the Main St stop to help determine the viability of the proposed business.
   1. This is also elaborated below at #6.
   2. Please find the code for this below
      1. <https://github.com/jabrown5/CTA-Ridership/blob/master/CTA_analyses.sas>

**Comparison to “L” Stations with Comparable Ridership Entry**

Stations with ridership entry most comparable to the Evanston Main Street “L” stop are listed below. These stations were found by calculating the overall weekday average for Evanston Main Street and then creating a range using the standard deviation for this stop. If other stops had a weekday average ridership falling within this range, they are included below.

|  |  |  |
| --- | --- | --- |
| **Obs** | **stationname** | **rides** |
| **2** | 51st | 1114.0797244 |
| **3** | California-Cermak | 1133.2204724 |
| **4** | Cermak-McCormick Pla | 1271.1022495 |
| **5** | Cicero-Forest Park | 1228.148376 |
| **6** | Damen-Cermak | 1109.2896161 |
| **7** | Francisco | 1266.6826686 |
| **8** | Main | 1192.4906496 |

Using the table below, weekly day of the week rides and corresponding standard deviations are listed. While ridership is roughly comparable, standard deviations for a handful of the stations is much greater than that of Evanston Main Street. Using these basic, preliminary descriptives, the 51st and Cicero-Forest Park “L” stations have similar ridership usage and more similar standard deviations indicating a steady ridership entry point.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Obs** | **stationname** | **dayOfWeek** | **rides** | **std\_deviation** |
| **14** | 51st | Monday | 1074.794155 | 170.66342877 |
| **15** | 51st | Tuesday | 1118.6953883 | 143.61550253 |
| **16** | 51st | Wednesday | 1127.8181818 | 144.02812363 |
| **17** | 51st | Thursday | 1118.0086741 | 144.0199744 |
| **18** | 51st | Friday | 1129.4384896 | 167.27067353 |
| **19** | California-Cermak | Monday | 1095.5031766 | 356.61433959 |
| **20** | California-Cermak | Tuesday | 1151.592233 | 357.38307146 |
| **21** | California-Cermak | Wednesday | 1153.5430303 | 358.40512875 |
| **22** | California-Cermak | Thursday | 1149.7211896 | 357.26366022 |
| **23** | California-Cermak | Friday | 1114.2959805 | 501.08278012 |
| **24** | Cermak-McCormick Pla | Monday | 1219.5789474 | 429.0975334 |
| **25** | Cermak-McCormick Pla | Tuesday | 1245.84 | 430.68378144 |
| **26** | Cermak-McCormick Pla | Wednesday | 1248.12 | 427.726295 |
| **27** | Cermak-McCormick Pla | Thursday | 1259.7628866 | 398.54746092 |
| **28** | Cermak-McCormick Pla | Friday | 1382.6391753 | 580.63740452 |
| **29** | Cicero-Forest Park | Monday | 1179.6696315 | 218.78942932 |
| **30** | Cicero-Forest Park | Tuesday | 1241.1089588 | 211.2134184 |
| **31** | Cicero-Forest Park | Wednesday | 1242.1116505 | 218.87719035 |
| **32** | Cicero-Forest Park | Thursday | 1243.3920596 | 218.04302416 |
| **33** | Cicero-Forest Park | Friday | 1232.6004872 | 222.46246315 |
| **34** | Damen-Cermak | Monday | 1059.1829733 | 392.80638947 |
| **35** | Damen-Cermak | Tuesday | 1121.2961165 | 402.39867029 |
| **36** | Damen-Cermak | Wednesday | 1127.3418182 | 407.45924152 |
| **37** | Damen-Cermak | Thursday | 1135.2180917 | 415.56813874 |
| **38** | Damen-Cermak | Friday | 1101.6443362 | 419.17917199 |
| **39** | Francisco | Monday | 1210.5857687 | 319.84125945 |
| **40** | Francisco | Tuesday | 1283.9793689 | 321.06585223 |
| **41** | Francisco | Wednesday | 1282.9004854 | 327.1617958 |
| **42** | Francisco | Thursday | 1290.0459057 | 329.70794398 |
| **43** | Francisco | Friday | 1263.8830694 | 340.60131327 |
| **44** | Main | Monday | 1138.3341804 | 101.39454537 |
| **45** | Main | Tuesday | 1206.4824242 | 88.445529252 |
| **46** | Main | Wednesday | 1186.3066667 | 92.267363018 |
| **47** | Main | Thursday | 1205.0086849 | 93.957086406 |
| **48** | Main | Friday | 1224.2691839 | 138.18684554 |



**Comparison to Evanston Davis Street**

While having greater ridership entry, the Evanston Davis Street stop was compared to the Evanston Main Street stop. This was because both are in Evanston and the Davis Street stop has a [Dunkin Donuts franchise within the CTA “L” station](https://www.golocal247.com/biz/dunkin-donuts/evanston-il/LOC143925507) and so could possibly inform the viability of a Dunkin Donuts (or similar) establishment at or near the Evanston Main Street stop.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Obs** | **stationname** | **dayOfWeek** | **rides** | **std\_deviation** |
| **5** | Davis | Monday | 3569.6493011 | 318.53065812 |
| **6** | Main | Monday | 1138.3341804 | 101.39454537 |
| **8** | Davis | Tuesday | 3738.217233 | 285.97801774 |
| **9** | Main | Tuesday | 1206.4824242 | 88.445529252 |
| **11** | Davis | Wednesday | 3718.6286408 | 299.7920753 |
| **12** | Main | Wednesday | 1186.3066667 | 92.267363018 |
| **14** | Davis | Thursday | 3799.7593052 | 287.59863618 |
| **15** | Main | Thursday | 1205.0086849 | 93.957086406 |
| **17** | Davis | Friday | 3991.6909976 | 442.25946209 |
| **18** | Main | Friday | 1224.2691839 | 138.18684554 |



* + Ideally, what other data would you gather or combine to learn even more?

I would propose several additional sources of data to supplement and add to existing knowledge gained through the Main St. Station entry data.

1. Busiest time of day for station entry
   1. This information could further inform establishment hours and staffing
   2. How to get this information?
      1. The CTA recently sent out an in-depth usage survey (I received this survey). If accessible, this data could be used to inform when most riders depart from Main St and when users disembark from this station.
      2. Visual observation
2. The type of rider
   1. Do riders use single ride passes? Or do most of the rides originate via regular users (with Ventra cards)?
3. Other public transportation options and ridership information
   1. Metra (Main Street Metra Station is just down the street)
      1. Metra boarding can be found at: <https://metrarail.com/about-metra/reports-documents/operations-ridership-data/station-level-data>
         1. Data is for Wednesday, April 2, 2014 only
         2. 1093 riders boarded at Evanston Main Street on this day

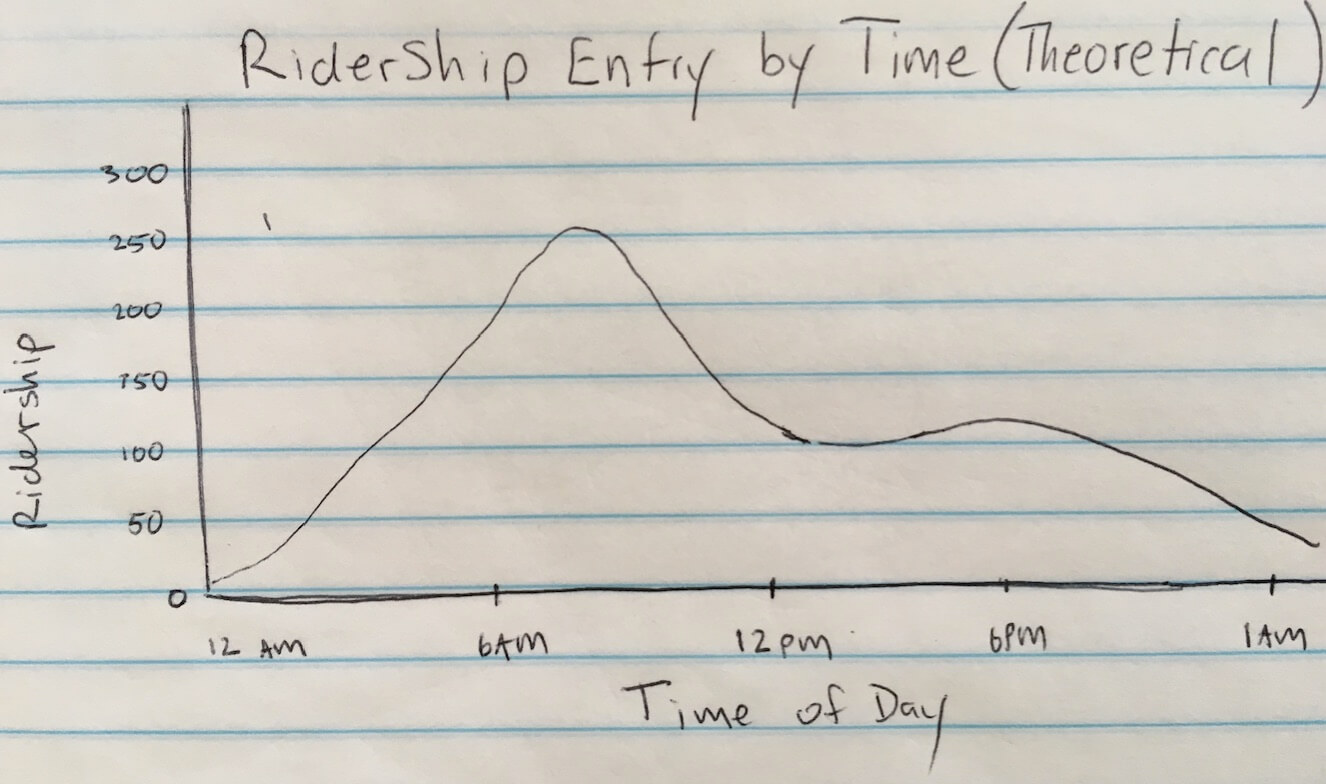
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Obs** | **stationname** | **date** | **rides** | **Metra rides** | **Total rides** |
| **1** | Davis | 04/02/2014 | 3914 | 2070 | 5984 |
| **2** | Main | 04/02/2014 | 1258 | 1093 | 2351 |

* 1. CTA bus options
     1. <http://www.transitchicago.com/assets/1/bus_stop_lists/205.htm>
     2. CTA bus ridership:
        1. 2012 data only: <https://data.cityofchicago.org/Transportation/CTA-Ridership-Avg-Weekday-Bus-Stop-Boardings-in-Oc/mq3i-nnqe>

1. Existing or competing business
   1. What breakfast food or coffee options exist?
      1. <https://www.google.com/maps/d/u/0/viewer?mid=19rVMj0ERAnxlr8NR7pnGoBqx5FY&hl=en_US&ll=42.03413867842476%2C-87.68001900000002&z=17>
2. Available real estate and associated costs
3. A more in-depth review of comparable neighborhood
   1. What does the comparison community look like? How is it similar or different to the Evanston Main St. community?
      1. Disposable income, etc.
      2. Census Data of Evanston Census tracts serviced by the Main St. “L”
         1. <https://censusreporter.org/profiles/14000US17031809900-census-tract-8099-cook-il/>
         2. <https://censusreporter.org/profiles/14000US17031809800-census-tract-8098-cook-il/>
         3. <https://censusreporter.org/profiles/14000US17031810100-census-tract-8101-cook-il/>
         4. <https://censusreporter.org/profiles/14000US17031810000-census-tract-8100-cook-il/>
   2. What do the businesses of these communities look like?
4. Can it be assumed that riders boarding are also returning to this stop at the end of the day? Specifically, can it be assumed that the majority of riders embarking from this station also live within its vicinity?
   * How would you want to see data presented, to make it actionable by you or others?

First, I’d like to keep in mind what the client might want. Second, I like to keep results mostly short and to the point. That being said, I also believe in presenting data a couple ways so there is ‘something for everyone’.

1. Brief text summaries to distill large amounts of information in an understandable format.
2. Tables.
   1. Specific to this project, I would include tables that outline usage by day of week (listed above), time of day (should the data become available), and mode (CTA, Metra, Bus)
   2. Tables including demographics of comparable neighborhoods
3. Visuals
   1. Weekday day usage (line graph listed above)
   2. Time of day (theoretical) line graph



* 1. Maps plotting existing business

**Why a Dunkin Donuts May Not Work at this Location**

* CTA policy prohibits food and drink on the CTA. If CTA customers abide by these rules, it may be unlikely the business can count on riders passing by to stop in and purchase a breakfast item. It is possible that customers may finish their purchases before boarding the train but that is an unknown variable. Conversely, food and drink is allowed on the Metra so perhaps these riders will still purchase breakfast or snack items.
* It’s yet unclear what Evanston’s policies toward franchises along the lake shore corridor are.
* There simply may not be a market for this type of business. While the numbers of riders may support a franchise location, it’s not clear if the consumers will actually patronize the business.
* See path diagram below for diagram of influences on the decision to open a new business and the type of business.

