

# Capstone Project – Final Report

*Clearly define a problem or an idea of your choice, where you would need to leverage the Foursquare location data to solve or execute. Remember that data science problems always target an audience and are meant to help a group of stakeholders solve a problem, so make sure that you explicitly describe your audience and why they would care about your problem.*

## 1. Introduction

### 1.1. Background

With summer rapidly approaching I am looking for opportunities to make extra money over the summer to pay for school. I would like to open a food truck that specializes in frozen treats. With the failure rate for food trucks being as high as 60%, I want to make sure I complete a sound business analysis. After all, it's about location, location, location.

### 1.2. Problem

I am not sure what location would be best for my food truck. I would like to find a location that is close to neighborhood parks, hoping that people will stop for a treat while they are walking their dog or exercising. Unfortunately, I am not willing to travel far from home, so the location will need to either be in my hometown of Nashville, TN or where I go to school in Knoxville, TN. In order to make sure I am not wasting my time I need to understand weather patterns that could discourage interest in frozen treats, as well as any COVID hotspots.

### 1.3. Audience

This analysis could be used by a number of people, but is mostly focused on owner/operators of food trucks in the central/eastern portion of Tennessee.

*Describe the data that you will be using to solve the problem or execute your idea. Remember that you will need to use the Foursquare location data to solve the problem or execute your idea. You can absolutely use other datasets in combination with the Foursquare location data. So make sure that you provide adequate explanation and discussion, with examples, of the data that you will be using, even if it is only Foursquare location data.*

## 2. Data Acquisition and Usage

### 2.1. Data Sources

I will use API to access venue data through Foursquare. The venue data will alert me to area parks, and I intend to use the tip count as a measure of popularity. Weather data will be scraped from the climate sections on the city Wikipedia pages. I will be focused on average temperature by month. COVID data will be extracted from Microsoft's COVID data file on GitHub. I am interested in total count for a specific region/lat long.

- Foursquare APIs - <https://developer.foursquare.com/docs/places-api/endpoints/>
- Weather data
  - Nashville - [https://en.wikipedia.org/wiki/Nashville,\\_Tennessee](https://en.wikipedia.org/wiki/Nashville,_Tennessee)
  - Knoxville: [https://en.wikipedia.org/wiki/Knoxville,\\_Tennessee](https://en.wikipedia.org/wiki/Knoxville,_Tennessee)
- COVID data - <https://github.com/microsoft/Bing-COVID-19-Data/tree/master/data>

## 2.2. Data Usage

The majority of data I need will come from the Foursquare APIs. This dataset will provide me with the location data about various parks in the area. I would also like to pull in historical weather information to understand what months are ideal for frozen treats. Lastly, I would like to try to avoid COVID hotspots, so I would like to pull in current COVID numbers by lat long. By merging these datasets together I hope to identify high traffic parks with warm temperatures that would encourage people to be outdoors.

Documentation of the Foursquare API:

### Regular API Endpoints

Method	Endpoint Group	Endpoint	Usage	Required?
GET	venues	<a href="#">search</a>	Search for Venues	—
GET	venues	<a href="#">explore</a>	Get Venue Recommendations	—
POST	venues	<a href="#">select</a>	Report Venue Selection	—
GET	venues	<a href="#">trending</a>	Get Trending Venues	—
GET	venues	<a href="#">suggestcompleti on</a>	Suggest Completions	—

Example of weather data from Wikipedia (Knoxville):

Climate data for Knoxville ( <b>McGhee Tyson Airport</b> ), 1981–2010 normals, <sup>[c]</sup> extremes 1871–present <sup>[d]</sup>													
Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Year
<b>Record high °F (°C)</b>	77.0 (25.0)	83.0 (28.3)	88.0 (31.1)	93.0 (33.9)	96.0 (35.6)	105.0 (40.6)	105.0 (40.6)	105.0 (40.6)	105.0 (40.6)	96.0 (35.6)	85.0 (29.4)	80.0 (26.7)	105.0 (40.6)
<b>Mean maximum °F (°C)</b>	67.2 (19.6)	71.3 (21.8)	79.1 (26.2)	84.6 (29.2)	87.9 (31.1)	92.8 (33.8)	95.2 (35.1)	94.6 (34.8)	91.1 (32.8)	83.7 (28.7)	76.6 (24.8)	68.3 (20.2)	96.1 (35.6)
<b>Average high °F (°C)</b>	47.3 (8.5)	52.3 (11.3)	61.4 (16.3)	70.3 (21.3)	78.1 (25.6)	85.4 (29.7)	88.2 (31.2)	87.8 (31.0)	81.8 (27.7)	71.2 (21.8)	60.4 (15.8)	49.8 (9.9)	69.6 (20.9)
<b>Average low °F (°C)</b>	29.2 (−1.6)	32.4 (0.2)	39.2 (4.0)	47.3 (8.5)	56.2 (13.4)	64.7 (18.2)	68.7 (20.4)	67.8 (19.9)	60.4 (15.8)	48.5 (9.2)	39.0 (3.9)	31.7 (−0.2)	48.8 (9.3)
<b>Mean minimum °F (°C)</b>	9.6 (−1.2)	14.2 (−9.9)	22.2 (−5.4)	30.8 (−0.7)	40.8 (4.9)	52.7 (11.5)	59.9 (15.5)	59.7 (15.4)	46.2 (7.9)	32.6 (0.3)	23.6 (−4.7)	14.8 (−9.6)	5.6 (−14.7)
<b>Record low °F (°C)</b>	−24 (−31)	−10 (−23)	1.0 (−17.2)	22.0 (−5.6)	32.0 (0.0)	42.0 (5.6)	49.0 (9.4)	49.0 (9.4)	35.0 (1.7)	24.0 (−4.4)	5.0 (−15.0)	−6 (−21)	−24 (−31)
<b>Average precipitation inches (mm)</b>	4.32 (110)	4.26 (108)	4.34 (110)	4.01 (102)	4.51 (115)	3.81 (97)	5.08 (129)	3.27 (83)	3.24 (82)	2.51 (64)	4.01 (102)	4.50 (114)	47.86 (1,216)
<b>Average snowfall inches (cm)</b>	2.7 (6.9)	1.6 (4.1)	0.9 (2.3)	0.5 (1.3)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	trace	0.8 (2.0)	6.5 (17)
<b>Average precipitation days (≥ 0.01 in)</b>	11.2	11.0	12.0	10.7	11.4	11.4	11.3	8.8	7.7	8.2	9.9	11.6	125.2

<b>Average snowy days (<math>\geq 0.1</math> in)</b>	1.5	1.2	0.6	0.1	0	0	0	0	0	0	0	0.8	4.2
<b>Average relative humidity (%)</b>	71.7	68.0	64.8	63.3	70.8	73.5	75.7	76.3	76.1	73.0	71.8	72.9	71.5
<b>Mean monthly sunshine hours</b>	135.8	145.3	208.9	256.6	287.2	291.1	287.3	278.0	232.3	217.2	151.7	122.5	2,613.9
<b>Percent possible sunshine</b>	44	48	56	65	66	67	65	67	62	62	49	40	59

Metadata for the Microsoft COVID dataset:

Column header	Description
ID	Unique identifier
Updated	Datetime in UTC
Confirmed	Confirmed case count for the region
ConfirmedChange	Change of confirmed case count from the previous day
Deaths	Death case count for the region
DeathsChange	Change of death count from the previous day
Recovered	Recovered count for the region
RecoveredChange	Change of recovered case counts from the previous day
Latitude	Latitude of the centroid of the region
Longitude	Longitude of the centroid of the region
ISO2	2 letter country code identifier

Column header	Description
ISO3	3 letter country code identifier
Country_Region	Country/region
AdminRegion1	Region within Country_region
AdminRegion2	Region within AdminRegion1