

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.

The data I am using is a csv file I obtained from <https://data.corkcity.ie>.

It is reasonably large measuring 4,975 rows by 34 columns. However of these the majority are unimportant to our investigation. As part of data preparation the following columns were kept and the rest dropped:

1. Application Number
  - a. As Primary Key
2. Latitude
3. Longitude
4. OneOffHouseFlag
  - a. This signals if the permission was for a one off house
5. NumberOfResidentialUnits
  - a. This signifies how many homes are being built
6. AreaOfSite
  - a. The total area of the property in km2
7. Decision
  - a. Whether the application was accepted for rejected

The new data frame looks like this:

ApplicationNumber	Latitude	Longitude	OneOffHouseKPI	NumResidentialUnits	AreaOfSite	Decision
832737	51.87518	-8.49607	No	0	0.043	Refused
832738	51.89802	-8.41491	No	0	0.074	Granted (Conditional)
832739	51.88336	-8.48256	No	0	0.02	Granted (Conditional)
832740	51.88769	-8.47358	Yes	1	0.042	Refused
832741	51.91723	-8.46982	No	0	0.85	Granted (Conditional)

In my Jupyter notebook it looks like this

```
[11]: planp = pd.read_csv('planningpermission2.csv')
      planp.head()
```

```
[11]:
```

	_id	ApplicationNumber	Latitude	Longitude	OneOffHouseKPI	NumResidentialUnits	AreaOfSite	Decision
0	3	832737	51.875175	-8.496067	No	0	0.043	Refused
1	4	832738	51.898020	-8.414907	No	0	0.074	Granted (Conditional)
2	5	832739	51.883356	-8.482564	No	0	0.020	Granted (Conditional)
3	6	832740	51.887688	-8.473576	Yes	1	0.042	Refused
4	7	832741	51.917226	-8.469818	No	0	0.850	Granted (Conditional)