

Use of Open data to prepare the tourist trip to MUC

Introduction and Background

Munich exudes Bavarian charm. Beer fanatics should head immediately to the Hofbräuhaus, a hops heaven that's been churning out the good stuff since 1589. The drinking is downright legendary during Oktoberfest, a celebration of local beers and German specialty foods. Emulate world-class athletes at the Olympiapark, where skating on the Olympic ice rink will make you feel like a champion. The promenade of Marienplatz is perfect for people watching and gawking at the Glockenspiele of City Hall¹. In Munich there are many museums, theaters and offcourse it is well known for a Beer Festival **Oktoberfest** with about 6 millions visitors per year.



Figure: Oktoberfest Venue (<https://en.wikipedia.org/wiki/Oktoberfest>)



Figure: Schleissheim Palace (https://en.wikipedia.org/wiki/Schleissheim_Palace)



Figure: Eisbach

Along with that, there are many parks and playgrounds and points of interest, which is worth to visit.

https://www.tripadvisor.com/Tourism-g187309-Munich_Upper_Bavaria_Bavaria-Vacations.html?fid=0c46b919-247a-458d-82d6-af73f4626e8e

Problem Description

When you planning tourist trip, you are often reading reports, suggestions and have expectations about points of interest to visit. As long as some point of interest is nearby, the priority can be suddenly changed. Sometimes you or your kids need WC, or wifi. Sometimes you can decide to go for lunch before visiting POI. Sometimes you spend less time at POI and still have several hours free. You want to discover what is nearby. Are there same sightseeing points or kids playground.

[Open data](https://en.wikipedia.org/wiki/Open_data) (https://en.wikipedia.org/wiki/Open_data) is the idea that some data should be freely available to everyone to use and republish as they wish, without restrictions from copyright, patents or other mechanisms of control. Many government and city authorities publish open data related to the country, city or region. While digging into these data sources it is possible to enrich normal tour guides and recommendations. In this project we will try to help different categories of tourist travellers to find places, like WiFi hotspots, Markets, Kids playgrounds, Sites around Isar river and Public WCs using Open Data and APIs like FourSquare.

Data sources

Foursquare API <https://developer.foursquare.com> (<https://developer.foursquare.com>).

POI at Isar River with coordinates <https://www.opengov-muenchen.de/dataset/points-of-interest-an-der-sudlichen-isar/resource/6fb56066-11b2-498d-ba22-cbcc85b660b4> (<https://www.opengov-muenchen.de/dataset/points-of-interest-an-der-sudlichen-isar/resource/6fb56066-11b2-498d-ba22-cbcc85b660b4>).

Public Wifi <https://www.opengov-muenchen.de/dataset/m-wlan-hotspots/resource/4ac5b616-ebbf-4857-ae79-43369d6c4c87> (<https://www.opengov-muenchen.de/dataset/m-wlan-hotspots/resource/4ac5b616-ebbf-4857-ae79-43369d6c4c87>).

Public Playgrounds <https://www.opengov-muenchen.de/dataset/oeffentliche-spielplaetze-muenchen/resource/845ce3bd-ea80-4623-b51d-a30680175c22> (<https://www.opengov-muenchen.de/dataset/oeffentliche-spielplaetze-muenchen/resource/845ce3bd-ea80-4623-b51d-a30680175c22>).

Oktoberfest data inkluding beer and fried chiken price over the years https://www.opengov-muenchen.de/dataset/oktoberfest/resource/e0f664cf-6dd9-4743-bd2b-81a8b18bd1d2?view_id=e619eb4dfc3b-44cc-88ea-0086d710ee1f (https://www.opengov-muenchen.de/dataset/oktoberfest/resource/e0f664cf-6dd9-4743-bd2b-81a8b18bd1d2?view_id=e619eb4dfc3b-44cc-88ea-0086d710ee1f).

Location of Markets <https://www.opengov-muenchen.de/dataset/maerkte/resource/f5608df8-43d3-4575-9bce-4fdb810bb516> (<https://www.opengov-muenchen.de/dataset/maerkte/resource/f5608df8-43d3-4575-9bce-4fdb810bb516>).

Museums Visitors <https://www.opengov-muenchen.de/dataset/monatszahlen-museen/resource/6c6a809e-91ee-4f3e-9268-a8b7bc38311c> (<https://www.opengov-muenchen.de/dataset/monatszahlen-museen/resource/6c6a809e-91ee-4f3e-9268-a8b7bc38311c>).

Public Toilets <https://www.opengov-muenchen.de/dataset/oefentliche-toiletten-der-landeshauptstadt-muenchen> (<https://www.opengov-muenchen.de/dataset/oefentliche-toiletten-der-landeshauptstadt-muenchen>).

Methodology and Results

Data normalization and enrichment

Before perform **data analysis** it's important to briefly look into data we have. We loaded data from various datasources into *pandas dataframes*. Here are the priliminary observations:

- majority of the data is presented in csv format
- no synchronisation even with splitters in the file. Three types of options was found ',' or ';' or '\t'.
- some datasources from open data in Munich have both csv and kml format.

- no synchronization among majority of field names
- some data sources, like information about museums, doesn't have coordinates or even addresses.

```
In [54]: 1 # Museums_Visitors
2 # https://www.opengov-muenchen.de/dataset/monatszahlen-museen/resource/6c6a809e-91ee-4f3e-9268-a8b7bc38311c
3 Museums_Visitors_URL = 'https://www.opengov-muenchen.de/dataset/bfb4a286-bea5-4bfe-82ce-b9bd354284a5/resource/6c6
4
5 museums_data = pd.read_csv(Museums_Visitors_URL)
6 print(len(museums_data))
7 museums_data.head()
8
9
```

2005

ZAHL	AUSPRAEGUNG	JAHR	MONAT	WERT	VORJAHRESWERT	VERAEND_VORMONAT_PROZENT	VERAEND_VORJAHRESMONAT_PROZENT	ZWO
1	Besucher/innen	Alte Pinakothek	2016	201612	15503.0	15344.0	11.89	1.04
1	Besucher/innen	Alte Pinakothek	2016	201611	13856.0	13584.0	-20.81	2.00
1	Besucher/innen	Alte Pinakothek	2016	201610	17498.0	22501.0	47.40	-22.23
1	Besucher/innen	Alte Pinakothek	2016	201609	11871.0	18027.0	-35.57	-34.15
1	Besucher/innen	Alte Pinakothek	2016	201608	18425.0	23021.0	-4.83	-19.96

Figure: Information about museums visitors without coordinates or addresses

Data sources overview

For the **central point on the maps**, address *Karlsplatz 1, 80335 Munich* were choosed using *Foursquare API*, the following coordinates were retrieved and used to center the maps for visual analysis: *latitude: 48.1400615 longitude: 11.5670315*.

POI at Isar River with

The data was downloaded using this URL: `_Isar_POIURL = 'https://www.opengov-muenchen.de/dataset/fc28edb7-0e69-47b2-814e-3fa18c68f94f/resource/6fb56066-11b2-498d-ba22-cbcc85b660b4/download/pois-an-der-isar-2016-08-09.csv'` (<https://www.opengov-muenchen.de/dataset/fc28edb7-0e69-47b2-814e-3fa18c68f94f/resource/6fb56066-11b2-498d-ba22-cbcc85b660b4/download/pois-an-der-isar-2016-08-09.csv>). The sample of the data provided below:

```
5 print(len(isar_data))
6 isar_data.head()
```

44

type	title	latitude	longitude	description	photo_url	photo_credit	url_webs
bruecke	Braunauer Eisenbahnbrücke	48.118377	11.561661	Wurde für die Eisenbahnlinie zwischen München ...	https://res.cloudinary.com/muenchen/isar-map/i...	Baureferat	http://orte.muenchen.de/1255494.html
bruecke	Brudermühlbrücke	48.112607	11.560013	Entstand durch Versetzung der alten Wittelsbac...	https://res.cloudinary.com/muenchen/isar-map/i...	Baureferat	http://orte.muenchen.de/1255509.html
bruecke	Corneliusbrücke	48.128557	11.580125	Wurde entweder nach dem Maler und Direktor der...	https://res.cloudinary.com/muenchen/isar-map/i...	Baureferat	https://www.muenchen.de/sehenswuerdigkeiten/c...
bruecke	Flauchersteg	48.104606	11.553159	Wurde 1914 zu einem soliden, von Betonpfeilern...	https://res.cloudinary.com/muenchen/isar-map/i...	Baureferat	http://orte.muenchen.de/1255470.html
bruecke	Föhringer Eisenbahnbrücke	48.183747	11.631017	War schon beim Bau 1907 mit einem Gleis auf de...	https://res.cloudinary.com/muenchen/isar-map/i...	Baureferat	N

Figure: The sample of the data with description of Points of Interest along Isar river

The field *title* gives the name of POI, coordinates of POIs provided in this data source.

Public WiFi

The data was downloaded using this URL: _Public_WIFIURL = '<https://www.opengov-muenchen.de/dataset/2f256473-423d-4c3d-846c-09977aa31205/resource/4ac5b616-ebbf-4857-ae79-43369d6c4c87/download/m-wlan2015-01-21.csv>''. The sample of the data provided below:

```

6 print(len(wifi_data))
7 wifi_data.head()
8
21

```

	Platz	Acess Point	latitude	longitude
0	Odeonsplatz	AP_1	48.142361	11.577892
1	Odeonsplatz	AP_2	48.142429	11.577247
2	Odeonsplatz	AP_4	48.142609	11.578021
3	Sendlinger Tor	AP_1	48.133764	11.566767
4	Sendlinger Tor	AP_2	48.133906	11.567297

Figure: The sample of the data with location of public WiFi Areas

In the field *Platz* it is possible to find hint for location, coordinates of points provided in this data source.

Public playgrounds

The data was downloaded using this URL: _Public_PlaygroundsURL = 'https://www.opengov-muenchen.de/dataset/0760ce3a-fef8-43e4-888f-8cc92fdf56de/resource/845ce3bd-ea80-4623-b51da30680175c22/download/spielplaetze_muenchen_hohne_leerespalten2016-06-13.csv'. The sample of the data provided below:

	adress_id	bezeichnung	latitude	longitude	service_id	service_kfz_parkplaetze_fuer_behinderte	service_hilfen_fuer_gehbehinderte	service_hilfen_fuer_sehbehinderte
0	10100152	Spielplatz "Schwere-Reiter-Str. Ecke Ackermann..."	48.163481	11.552318	10100153		Nein	Nein
1	10101944	Fitness Parcours "Südpark"	48.104765	11.507671	10101945		Nein	Nein
2	10101948	Fitness Parcours "Feldmochinger See"	48.215524	11.513373	10101949		Nein	Nein
3	10101950	Fitness Parcours "Merseburger Straße"	48.187803	11.516682	10101951		Nein	Nein
4	10101952	Fitness Parcours "Hohenbergplatz"	48.136959	11.507236	10101953		Nein	Nein

Figure: The sample of the data with location of playgrounds

In the field *title* we can find a name of the place. It is possible to find hint for location in the field *description*, coordinates of points provided in this data source.

The number of playgrounds is over seven hundreds and it's a huge number. It could be pure playground for kids or places, which are suitable for fitness exercises for adults and young generation. This kind of places are very important for travellers, who want to keep their physical shape.

Oktoberfest data

The data was downloaded using this URL: _OktoberfestURL = '<https://www.opengov-muenchen.de/dataset/8d6c8251-7956-4f92-8c96-f79106aab828/resource/e0f664cf-6dd9-4743-bd2b-81a8b18bd1d2/download/oktoberfestgesamt19852018.csv>' (<https://www.opengov-muenchen.de/dataset/8d6c8251-7956-4f92-8c96-f79106aab828/resource/e0f664cf-6dd9-4743-bd2b-81a8b18bd1d2/download/oktoberfestgesamt19852018.csv>). The sample of the data provided below:

```
: 1 # Oktoberfest data inkluding beer and fried chicken price over the years
 2 # https://www.opengov-muenchen.de/dataset/oktoberfest/resource/e0f664cf-6dd9
 3
 4 Oktoberfest_URL = 'https://www.opengov-muenchen.de/dataset/8d6c8251-7956-4f9
 5 oktoberfest_data = pd.read_csv(Oktoberfest_URL)
 6 print(len(oktoberfest_data))
 7 oktoberfest_data.head()
 8
 9
```

34

	jahr	dauer	besucher_gesamt	besucher_tag	bier_preis	bier_konsum	hendl_preis	hendl_konsum
0	1985	16	7.1	444	3.20	54541	4.77	629520
1	1986	16	6.7	419	3.30	53807	3.92	698137
2	1987	16	6.5	406	3.37	51842	3.98	732859
3	1988	16	5.7	356	3.45	50951	4.19	720139
4	1989	16	6.2	388	3.60	51241	4.22	775674

Figure: The sample of the data with statistics about Oktoberfest Beer festival since 1985 till 2018

No information about location is provided in the file, but data like _number of visitors per year (besucher_gesamt), per day (besucher_tag) in thousands, how long was the festival in particular year (dauer) and most important - information about the price of the beer (bier_preis) and roasted chicken (hendl_preis) in euros and how much beer were consumed and how many chickens were eaten. We used address *Theresienwiese, Bavariaring, 80336 München* and Foursquare API, to determine the following coordinates: *latitude: 48.1358538 longitude: 11.5495032*. These coordinates suggest, that the venue is walking distance from city center.

Location of Markets

The data was downloaded using this URL: _ILocation_Of_MarketsURL = '<https://www.opengov-muenchen.de/dataset/04eda73a-173f-43c4-abb4-3b3166c3f356/resource/f5608df8-43d3-4575-9bce-4fdb810bb516/download/141212lhmaerktealleattribute.csv>' (<https://www.opengov-muenchen.de/dataset/04eda73a-173f-43c4-abb4-3b3166c3f356/resource/f5608df8-43d3-4575-9bce-4fdb810bb516/download/141212lhmaerktealleattribute.csv>). The sample of the data provided below:

```

5 print(len(markets_data))
6 markets_data.head()

```

51

	xcoord	ycoord	fid	ogr_fid	etyp	ebene	faktor	winkel	name	inhalt	adresse	link
0	4465788.32	5331456.22	0	34751	SY	309	1	90	WMARKT	Untersendling	Margaretenplatz	http://www.muenchen.de/rathaus/Stadtverwaltung... Di
1	4464188.72	5326632.74	1	34752	SY	309	1	90	WMARKT	Solln	Fellerer Platz	http://www.muenchen.de/rathaus/Stadtverwaltung... Di
2	4457992.57	5334652.99	2	34753	SY	309	1	90	WMARKT	Neuaubing	Mainaustraße 67-73	http://www.muenchen.de/rathaus/Stadtverwaltung... Di
3	4469672.14	5330271.84	3	34754	SY	309	1	90	WMARKT	Giesing	Perlacher Straße 111 - 119	http://www.muenchen.de/rathaus/Stadtverwaltung... Di
4	4461389.89	5328035.46	4	34755	SY	309	1	90	WMARKT	Fürstenried- West	Graubündener Straße (Schweizer Platz)	http://www.muenchen.de/rathaus/Stadtverwaltung... Di

Figure: The sample of the data about Markets in Munich

The field *Name* doesn't give the information, that everybody expected. Field *inhalt* gives the distinct of the city, where market is located. Coordinates of POIs provided in this data source. The majority of markets give opportunity to buy local tasty and cheap food. This information is widely used by experienced travellers.

Museums visitors

The data was downloaded using this URL: `_IMuseums_VisitorsURL = 'https://www.opengov-muenchen.de/dataset/bfb4a286-bea5-4bfe-82ce-b9bd354284a5/resource/6c6a809e-91ee-4f3e-9268-a8b7bc38311c/download/monatszahlen1810museen.csv'`. The sample of the data was provided in beginning of the section. Since these data doesn't contain coordinates, it requires enrichment. *Foursquare* API was used to determine the location of museums. The result is shown below:

```

10 # Collecting coordinates
11 for index, row in museum_addresses.iterrows():
12
13     address = row['POI_Address']
14     location = geolocator.geocode(address)
15     latitude = location.latitude
16     longitude = location.longitude
17     print(row['POI_Name'], ":", latitude, longitude)
18     # time.sleep(7)
19     museum_addresses.at[index, "latitude"] = latitude
20     museum_addresses.at[index, "longitude"] = longitude
21
22
23
24 museum_addresses
25 poi_locations = museum_addresses.copy()
26 poi_locations

```

Alte Pinakothek : 48.1482861 11.569970904087043
 Bayerisches Nationalmuseum : 48.1433855 11.5911498
 Deutsches Museum - Museumsinsel : 48.13003865 11.582888488551909
 Deutsches Museum - Verkehrszentrum : 48.1325395 11.5429532
 Münchner Stadtmuseum : 48.1351603 11.572607585081988
 Museum Brandhorst : 48.1476318 11.57385874695941
 Museum Mensch und Natur : 48.15908585 11.51166615
 Neue Pinakothek : 48.1499278 11.570937445883057
 Pinakothek der Moderne : 48.14716025 11.572227101401982
 Schackgalerie : 48.14250955 11.593434725593049
 Städtische Galerie im Lenbachhaus : 48.1466458 11.5633122
 48.1358538 11.5495032

Figure: The sample of the enriched data with coordinates of museums

Public toilets

The data was downloaded using this URL: _Public_WIFI_URL = 'Public_ToiletsURL = '<https://www.opengov-muenchen.de/dataset/6d8552a7-328c-48a2-ab37-5f9caa2bfbdb/resource/cc327030-a5e4-4a27-8a99-aaa8a5ab6200/download/oeffentliche-toilettenmuenden2016-06-28.csv> (<https://www.opengov-muenchen.de/dataset/6d8552a7-328c-48a2-ab37-5f9caa2bfbdb/resource/cc327030-a5e4-4a27-8a99-aaa8a5ab6200/download/oeffentliche-toilettenmuenden2016-06-28.csv>)'. The sample of the data provided below:

address_organisation	address_organisationsbereich	bezeichnung	address_strasse	address_hausnummer	address_postleitzahl	address_ort	latitude	longit
Landeshauptstadt München	Referat für Gesundheit und Umwelt	Neuer Südfriedhof	Hochaeckerstr.	90	81737.0	München	48.100595	11.6224
Landeshauptstadt München	MTG	Öffentliche Toilette "Lorettoplatz"	Lorettoplatz	NaN	NaN	NaN	48.103475	11.4890
Landeshauptstadt München	Kommunalreferat	Öffentliche Toilette "Bavariring 5"	Bavariring	5	NaN	NaN	48.136390	11.5498
Landeshauptstadt München	Baureferat	Öffentliche Toilettenkabine Isarauen	westl. Marienklausenbrücke	NaN	NaN	NaN	48.092270	11.5478
Landeshauptstadt München	Baureferat	Öffentliche Toilettenkabine Isarauen	Floßlände	NaN	NaN	NaN	48.095602	11.5448

Figure: The sample of the data with location of public WCs Overall it is 134 locations. The file contains many fields. Address is splitted into separate fields like _Street (address_strasse), house number(address_hausnummer), postal index (addresspostleitzahl). Coordinates of points provided in this data source. The name can be taken from the following field _serviceanzeigename. The source contains WCs extra features like, does **douche** is present (duschen) and does it includes **Baby changing rooms** (wickelraeume). These features will be discovered in the *_data analysis* section of the document.

Combining information into single data structure which is suitable for visualization on the map

For a *Visual Data Analysis* we consider to use *Folium library* and display *POIs* on the *Map*. To do this we are importing *Names*, *Addresses*, *location data*, *Categories* of *POIs*. The *Categories* based on the datasources and Include '*Museum*', '*Festival*', '*POIs Near south Isar*', '*Public WiFi Acces Points*', '*Kids Playgrounds*', '*Markets*', '*Public WC*'. All information was loaded into single structure. The fragment of resulting data provided below:

	POI_Name	POI_Address	POI_Type	latitude	longitude
0	Alte Pinakothek	Barer Str. 27, 80333 München	Museum	48.148286	11.569971
1	Bayerisches Nationalmuseum	Prinzregentenstraße 3, 80538 München	Museum	48.143386	11.591150
2	Deutsches Museum - Museumsinsel	Museumsinsel 1, 80538 München	Museum	48.130039	11.582888
3	Deutsches Museum - Verkehrszentrum	Am Bavariapark 5, 80339 München	Museum	48.132540	11.542953
4	Münchner Stadtmuseum	Sankt-Jakobs-Platz 1, 80331 München	Museum	48.135160	11.572608
5	Museum Brandhorst	Theresienstraße 35a, 80333 München	Museum	48.147632	11.573859
6	Museum Mensch und Natur	Schloß Nymphenburg, 80638 München	Museum	48.159086	11.511666
7	Neue Pinakothek	Barer Str. 29, 80799 München	Museum	48.149928	11.570937
8	Pinakothek der Moderne	Barer Str. 40, 80333 München	Museum	48.147160	11.572227
9	Schackgalerie	Prinzregentenstraße 9, 80538 München	Museum	48.142510	11.593435
10	Städtische Galerie im Lenbachhaus	Luisenstraße 33, 80333 München	Museum	48.146646	11.563312
0	Oktoberfest	Theresienwiese, Bavariaring, 80336 München	Festival	48.135854	11.549503

Figure: The sample of the aggregated POIs data with categories and coordinates

Data Analysis

In *Data Analysis* section we consider to apply several approaches to discover the important and useful data for travellers. This section is splitted into four sections focused on different data mining techniques, inscluding simple, visual, prediction of prices using linear regression and statistical analysis.

Simple analysis: checking for Public WCs with special facilities

For simple analysis we choose an open data source for Public WC located in Munich. We will try to discover auxiliary features of WCs.

Public WCs with Douches

Some travellers jumping into the city walk right after the train or plain has arrived. If the weather conditions are hot or due to some other reasons, they may want to take a shower berfore hotel check-in or due to a single day trip. The question we answer here, does they have a chance to make it. The figure below says, unfortunately not. Based on the information, provided in Open Data, there is no single WC with these facilities

```

1 # Toilets with douche
2 wc_douche_data= wc_data[wc_data.duschen == True].copy()
3 print(len(wc_data))
4 print(len(wc_douche_data))
5 wc_douche_data.head()
6
7 # Probably No chance :-(
```

134

0

Figure: Zero Public WCs with shower found in the open data

Public WCs with Baby Changing Rooms

Baby changing rooms is an another important factor for travellers with kids. In Open data we discovered three public WCs with baby changing rooms.

```

In [51]: 1 # baby changing rooms
2
3
4 wc_bcr_data= wc_data[wc_data.wickelraeume == True].copy()
5 print(len(wc_data))
6 print(len(wc_bcr_data))
7 wc_bcr_data.head()
8
```

134

3

Out[51]:

	address_id	address_organisation	address_organisationsbereich	bezeichnung	adre
82	1065675	Landeshauptstadt München		MTG	Öffentliche Toilette "Marienplatz (im S- u. U-...)
83	1065677	Landeshauptstadt München		MTG	Öffentliche Toilette "Marienplatz Neues Rathau...
88	1065687	Landeshauptstadt München		MTG	Öffentliche Toilette "Münchner Freiheit (im U-...)

Figure: Three Public WCs with baby changing rooms found in the open data

Visual Analysis

Human brain naturally processing visual information way better when text. According to <http://www.t-sciences.com/news/humans-process-visual-data-better> (<http://www.t-sciences.com/news/humans-process-visual-data-better>): "Visualization works from a human perspective because we respond to and

process visual data better than any other type of data. In fact, the human brain processes images 60,000 times faster than text, and 90 percent of information transmitted to the brain is visual. Since we are visual by nature, we can use this skill to enhance data processing and organizational effectiveness." That's why in this project we visualizing open data on the map to see which points of interest located where and how they far away from each other. We used two types of map visualization with *colored circles* and *icons* in folium library. This gives an opportunity to compare how easy to find interesting locations using these two approaches. To avoid repeatable code, we compiled all POIs data into single file and displayed information on the map together. Figures below show both variants of visualization with *colored circles* and *icons*:

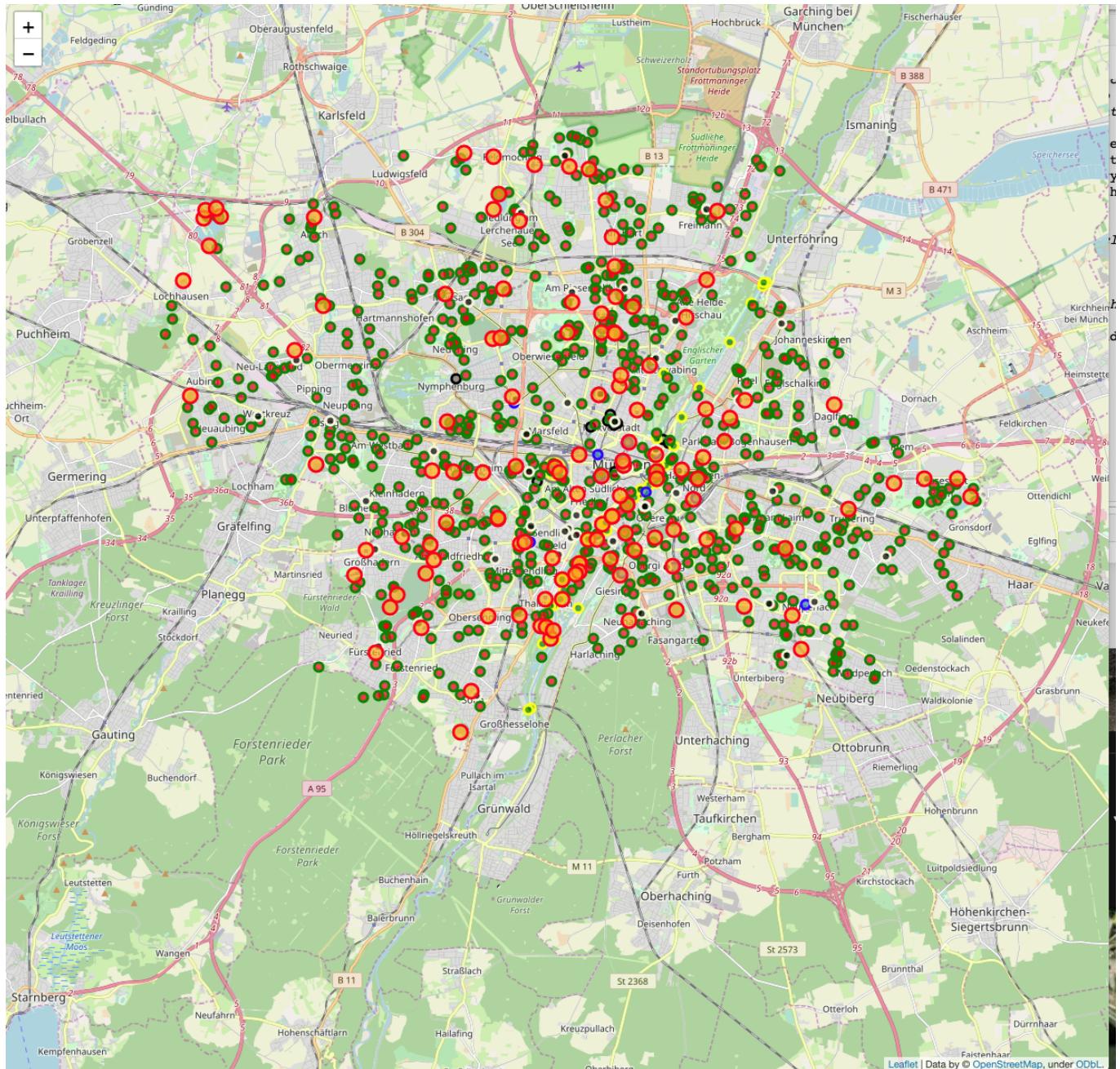


Figure: POI visualization with *colored circles* on the Map

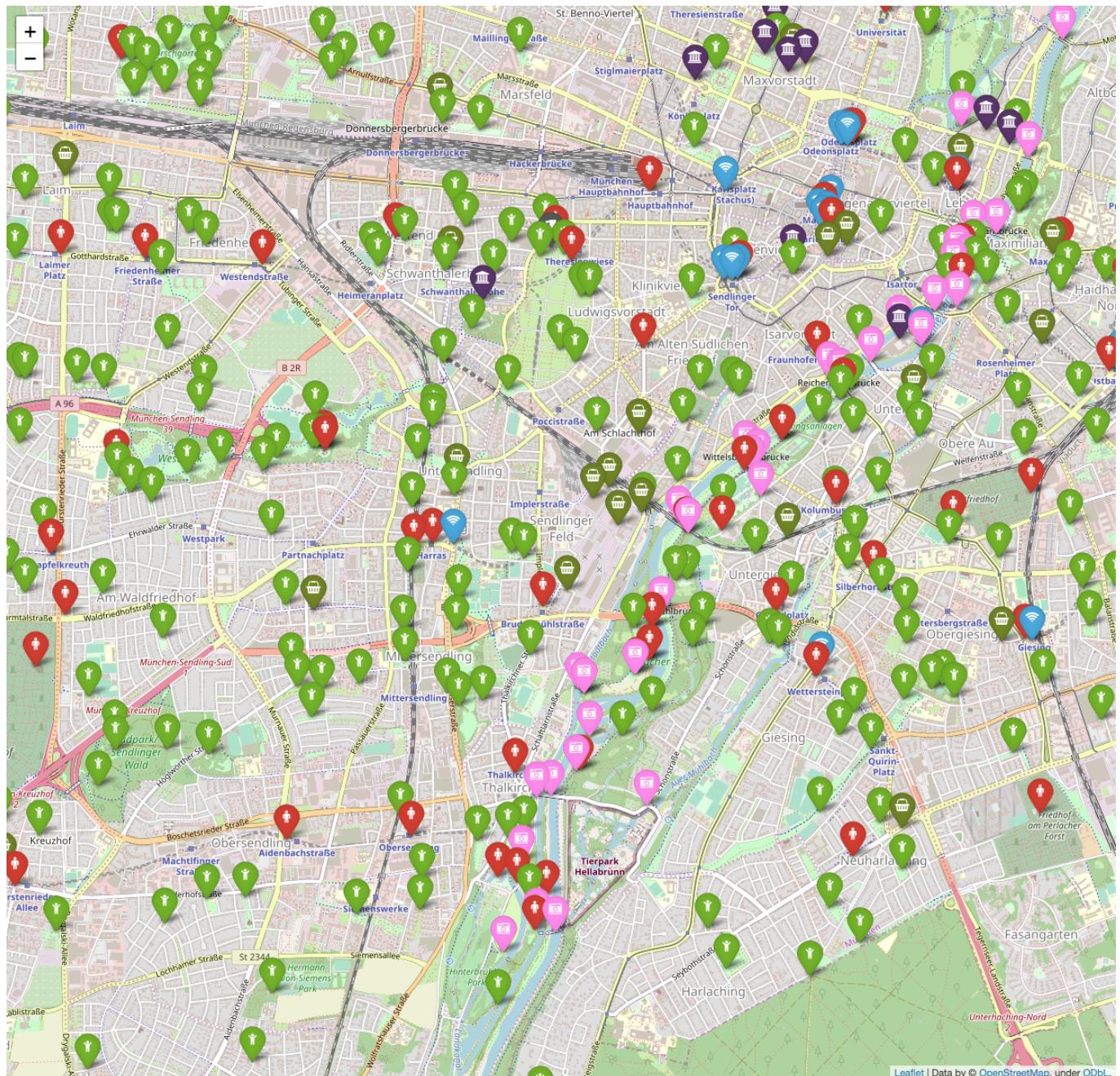


Figure: POI visualization with icons on the Map Both maps available on the github repository with names '_munich_withcircles.html' and '_munich_withicons.html'.

On the map with *colored circles* we used two colours for countour and fill to separate different types of POIs.

```

1 #Assigning Colors for different points
2
3 POIs_Types = POIs.POI_Type.unique()
4
5 color_scheme= pd.DataFrame({'POI_Type': ['Museum', 'Festival', 'POIs Near south Isar', 'Public WiFi Acces Points',
6 'POI_color': ['black', 'gray', 'yellow', 'blue', 'green', 'beige','red'],
7 'POI_fill': ['grey', 'yellow', 'green', '#3186cc', 'red', 'black','orange']}
8
9
10
11

```

Figure: Mapping of colors for different types of POIs

Another approach we use to make the map with *colored circles* more visual and interactive. We highlighted Oktoberfest venue and Public WCs with bigger circles.

```

7   # Making different color for different types and highlight the most important spots
8   radius = 5
9   # Oktoberfest must be visible
10  if pname == 'Oktoberfest' :
11      radius = radius*3
12  # Along with toilets
13  if ptype == 'Public WC' :
14      radius = radius*1.5
15

```

Figure: Bigger circles are used for Oktoberfest venue and Public WCs

As the result these venue and important POIs are more visible on the map. To make both maps more informative and interactive, we assigned *Hints* to every POI on the map. This gives opportunity to get way more information about particular POI.



Figure: Result with increased colored circles and hints on the Map

Bot visualisations are interactive and allow to find necessary places or places nearby. They are helping travellers in many scenarios suitable for different travellers groups. These scenarios can include but not limited to:

- "I just exited from railroad station, where is the nearest free WiFi"
- "My kids a bit tired with sightseeing, where is the nearest playground"
- "We just finished with Museum, where is the nearest market to try something local"
- "Oh, it's the end of September, where is Oktoberfest venue"
- "Too much beer inside, where is the nearest public..."

Figure below shows how questions from these scenarios can be easily solved, by navigating the map.

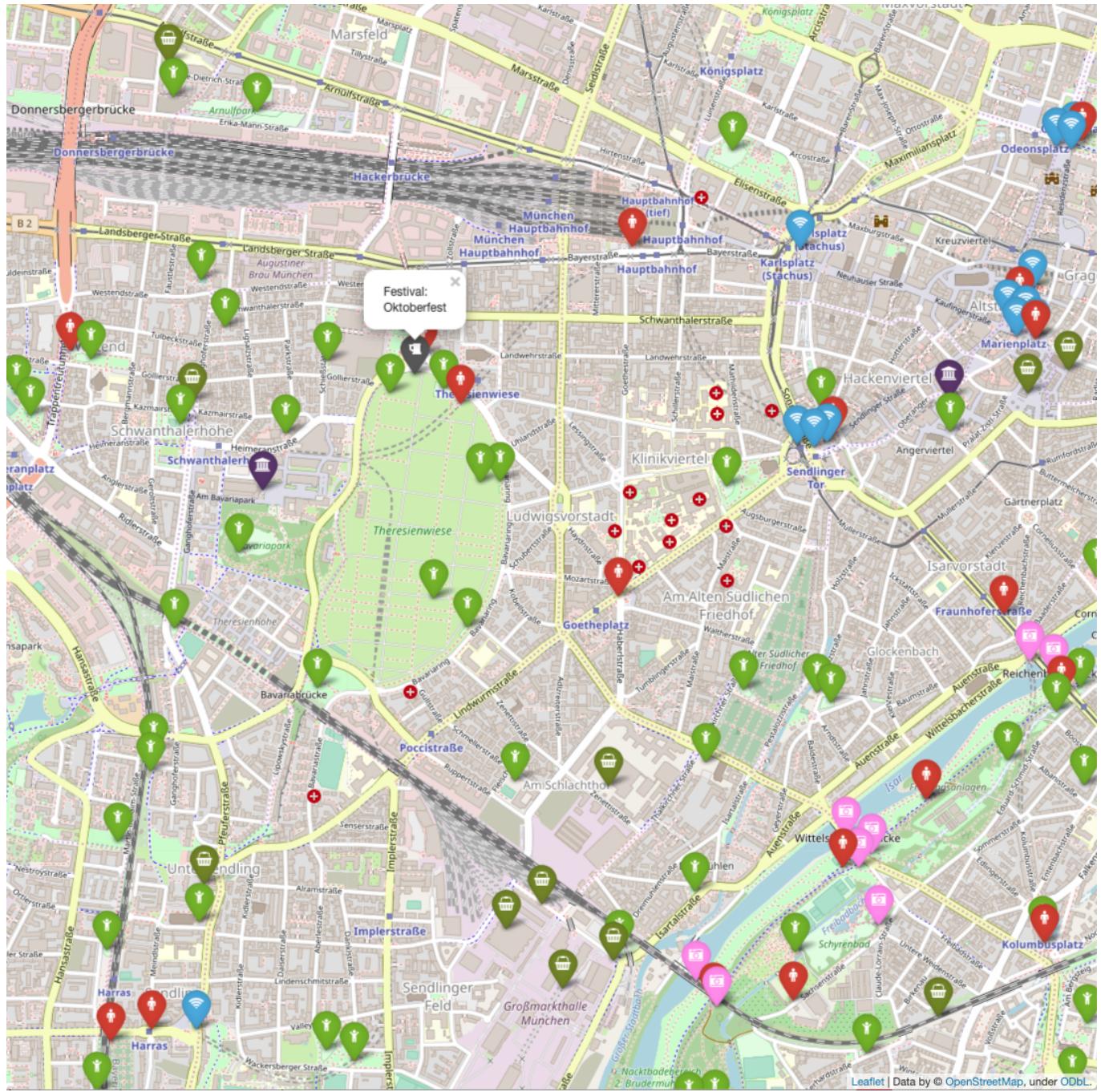


Figure: Map with **icons** and **hints** can helps in many scenarios

Beer and Chicken Price Prediction at Oktoberfert

Oktoberfest is one of the most visited events during the year. The number of visitors usually exceed 6 millions every year. This event gives visitors many new impressions, but it's not surprising, that beer and food cost money. One of the sources of Open Data includes the prices for beer and chicken at the festival venue till 2018. In this section we will try to predict the prices for the next several years using *sklearn linear regression*. We use the data from 1985-2018 to train the model

```

2 columns = ["Year", "Mass Bier Price", "Roasted Chicken Price"]
3 py = [[2019],[2020],[2021],[2022],[2023],[2024],[2025]]
4 X = np.array(oktoberfest_data.jahr.values.reshape(-1,1))
5 y = np.array(oktoberfest_data.bier_preis.values.reshape(-1,1))
6 z = np.array(oktoberfest_data.hendl_preis.values.reshape(-1,1))
7 regb = LinearRegression().fit(X, y)
8 regb.score(X, y)
9
10
11 beer=pd.DataFrame(regb.predict(py))
12
13 regc = LinearRegression().fit(X, z)
14 regc.score(X, z)
15 year= pd.DataFrame(py)
16 chicken = pd.DataFrame(regc.predict(py))
17
18 #Saving Predictions
19 predictions=pd.DataFrame(columns= columns)
20 predictions["Year"] = year[0]
21 predictions["Mass Bier Price"] = beer[0]
22 predictions["Roasted Chicken Price"] = chicken[0]
23 predictions.head(10)
24
25 # Normalizing historical data
26 historical = pd.DataFrame(columns= columns)
27 historical["Year"] = oktoberfest_data.jahr.copy()
28 historical["Mass Bier Price"] = oktoberfest_data.bier_preis.copy()
29 historical["Roasted Chicken Price"] = oktoberfest_data.hendl_preis.copy()
30 historical.tail(10)
31

```

Figure: Creating model for beer and chicken price prediction

```

31
32 prices = pd.concat([historical, predictions])
33 prices.tail(10)
34 plt.title("Beer and Chicken prices Historical till 2018 plus Predicted")
35 plt.xlabel("Year")
36 plt.ylabel("Price (Euro)")
37 plt.plot(prices["Year"], prices["Mass Bier Price"], 'r', label="Beer Price")
38 plt.plot(prices["Year"], prices["Roasted Chicken Price"], 'b', label="Chicken Price")
39 plt.legend()
40
41 plt.show()

```

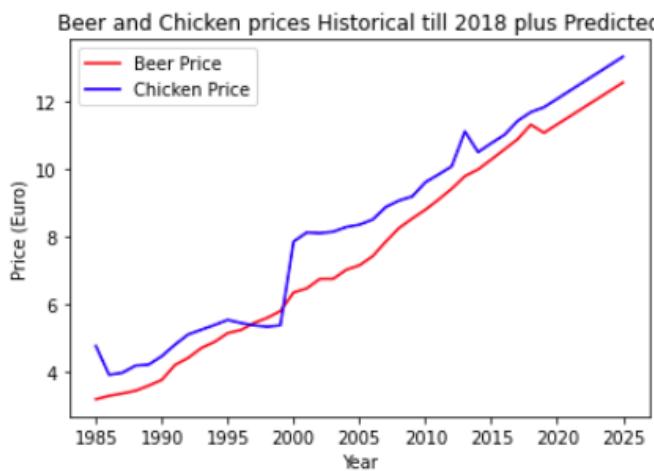


Figure: Visualizing the results for beer and chicken price prediction

As we can see on the visual, the prices for both, beer and chicken were increased almost every year. As shown on the figure below, the expected price for the beer is about 12.5 Euro and for chicken about 13.30 Euro in 2025. That means, the early people visit Oktoberfest, the cheaper beer they get.

	Year	Mass Bier Price	Roasted Chicken Price
0	2019	11.058984	11.812620
1	2020	11.306455	12.061459
2	2021	11.553927	12.310298
3	2022	11.801398	12.559137
4	2023	12.048869	12.807976
5	2024	12.296341	13.056814
6	2025	12.543812	13.305653

Figure: Numerical results for beer and chicken price predicion

Statistical analysis of most busy and easy months in Museums

Munich and surroundings is one of the top tourist destinaitons in Europe. Thare are many museums, castles, historical buildings here. But what travellers expect: come to museums, not spend several hours in the ques. In this section we esteemating total numberer of visitors in the museums in particular month, found in *OpenData* datasource. Based on this information we can determine most busy on most easy month in the particular museum. To do this, initially we grouping existing data by museum name and month and calculating the sum of visitors in particular month in particular museum over the years.

```

1 # Grouping and calculating how many visitors in every museum in total in particular month
2 ## Since month is mixed with year, removing year from month
3 museums_data[ "month" ] = museums_data[ "MONAT" ] - museums_data[ "JAHR" ]*100
4 museums_data[ "month" ]
5 columns =[ "month" , "museum" , "visitors" ]
6 mvisitors=pd.DataFrame(columns=columns)
7
8 mvisitors[ "month" ] = museums_data[ "month" ].copy()
9 mvisitors[ "museum" ] = museums_data[ "AUSPRAEGUNG" ].copy()
10 mvisitors[ "visitors" ] =museums_data[ "WERT" ].copy()
11
12 #mvisitors[ "visitors" ] =
13 mvisitors=mvisitors.groupby( [ mvisitors[ "museum" ],mvisitors[ "month" ]]).sum().reset_index()
14
15
16 mvisitors.tail(3)
17
18
19
:
```

	museum	month	visitors
129	Städtische Galerie im Lenbachhaus	10	424150.0
130	Städtische Galerie im Lenbachhaus	11	430141.0
131	Städtische Galerie im Lenbachhaus	12	425534.0

Figure: Aggregation of data by museum name and month

After calculation of the number of visitors we looking for minimum and maximum numbers for every museum to predict Go and No-Go time. We can see that for **six out of ten museums** September predicted as the best time to visit.

```

6 # Collecting less busy months for every museum
7 mn=mvisitors.groupby( mvisitors[ "museum" ] )[ "visitors" ].transform('min') ==mvisitors [ "visitors" ]
8 busy_easy_museum_months = mvisitors[mn].copy()
9 busy_easy_museum_months.rename(columns={"month": "month_easy", "visitors": "visitors_min"}, inplace = True)
10
11 # Collecting most busy months for every museum
12 mx=mvisitors.groupby( mvisitors[ "museum" ] )[ "visitors" ].transform('max') ==mvisitors [ "visitors" ]
13 busy_easy_museum_months1 = mvisitors[mx].copy()
14 busy_easy_museum_months1.rename(columns={"month": "month_busy", "visitors": "visitors_max"}, inplace = True)
15
16 # Merging Data into single table
17 bm=pd.merge(busy_easy_museum_months,busy_easy_museum_months1)
18
19 bm

```

:1:

	museum	month_easy	visitors_min	month_busy	visitors_max
0	Alte Pinakothek	9	319819.0	8	474533.0
1	Bayerisches Nationalmuseum	9	109722.0	12	308245.0
2	Deutsches Museum - Museumsinsel	11	1100457.0	8	2330174.0
3	Deutsches Museum - Verkehrszentrum	9	67830.0	8	115807.0
4	Museum Brandhorst	4	84969.0	10	153702.0
5	Museum Mensch und Natur	12	84070.0	3	182075.0
6	Münchener Stadtmuseum	8	139913.0	3	290504.0
7	Neue Pinakothek	9	232910.0	8	394217.0
8	Pinakothek der Moderne	6	360013.0	10	685313.0
9	Schackgalerie	9	8174.0	10	23256.0
10	Städtische Galerie im Lenbachhaus	9	230150.0	1	484913.0

Figure: Resulting data with most busy and easy month in particular museum

Discussion

Open data provided by Munich can't create fool picture for the traveller, but can definitely contribute to positive travel experience. As expected there were several issues with the data:

- data is outdated, now it's March 2020 and some data is several years old
- data formats are not synchronized. There were three different delimiters in the data sources we analyzed
- data fields are not synchronized, since original sources of data are different, we have many names for the same fields, and sometimes names don't match the data or contain the same value in every column for the name.
- data missing important fields. That means, that every dataset has to be manually reviewed before processing. During processing, particular fields with different names should be matched to do correlations, third party sources should be used to enrich the data. In general, these things shouldn't surprise you while you're digging into Open Data Topic. Despite that, in the short period of time it is possible to find a lot of interesting data for data scientist in Open data if you want to dig into particular application field. In this project we focused on demonstrating how a data scientist can mine meaningful information from open data with focus on the travellers as a target audience. There are still many things that could be done for improvement, but considering approximately 20 hours in total for programming, whitepaper and presentation we should know where to stop and give opportunity for others to follow up.

Conclusion

In this project we used *Open Data* published by Munich to mine the information for tourist trip to this wonderful city. We normalized the data and enriched datasets using *Foursquare API*. After that we used several techniques, *simple data analysis* to determine extra features of Public WCs, *visualisation* to find different types of places, like kids playgrounds, markets, museums, sightseeing points, WiFi hotspots on the interactive Munich map that we created. We used *linear regression from sklearn library* for beer and chicken price prediction and *statistical analysis* to determine the best time to visit museums. All of these can help

travellers to improve their standart and unusual travel experience in this wonderful city. Hope that ideas of this project demonstrate several data science techniques and can help different categories of tourist

In []: