Project: Improving the City of Montevideo
Author: Jorge Vallego

INTRODUCTION

Target Audience

Government of Montevideo

Background

Montevideo is the capital of Uruguay and by large margin the biggest city in the country. It is located within its homonymous Department (note: a department is somewhat equivalent to a state in US).
For administrative purposes, the city is divided in eight Municipalities. The population of Montevideo is about 1.3 Million people in an area of 200 sq kms.

Problem

The Government of Montevideo is interested in improving the city and the quality of live of its inhabitants, and also make it more attractive for small and large business. Ideally they want to do as much of the three improvements without doing one at the expense of another.

The perceptions is that there are areas more in need of improvements than others and the satisfaction of the citizens might be a good indicator that can be used as a way to identify those areas.
As there recently has been an election and the results are now available, this poll can be used as and indicator of satisfaction; where the Government has obtained a low share of the votes in relation to the opposition, it would be taken as an area where more needs to be done for the satisfaction of its inhabitants.

The idea of this project is then to compare areas, in terms of type and number of venues and services, and also include the poll results as an' indicator of citizen's satisfaction.

Benefits of this project

With the insight provided by this study, the Government of Montevideo will try and find ways to encourage business, institutions, etc. to move their venues, or open new ones, where there is a lack of them. This will not only be good for neighbourhoods and their inhabitants but will also bring business opportunities and reduce the use of commuting resources.

DATA

Sources

Raw data
The following publicly available information has been used to acquire the source data for this project.

For Municipalities data:

https://municipioe.montevideo.gub.uy/preguntas-frecuentes/donde-estan-ubicados-los-municipios-de-montevideo

For poll results:

https://eleccionesnacionales.corteelectoral.gub.uy/json/Exportaciones/Inf_D_Lema.xlsx

For connecting polls to municipalities:

https://registronacional.com/uruguay/elecciones/series-electorales-de-municipios-en-uruguay.htm
https://registronacional.com/uruguay/elecciones/circuitos-electorales-en-montevideo.htm

For venue information

Foursquare API: https://api.foursquare.com/v2/venues

Data Preparation

The above date required to be formatted and cleaned (for example series and municipalities are just text within a web page) and then merged with the result of the polls in order to be able to import it into a data frames.
The data sources that will be the basis for the analysis has been uploaded here:

https://github.com/DSS-TAJorge/Coursera_Capstone/blob/master/MunicipiosAddresses.xlsx

https://github.com/DSS-TAJorge/Coursera_Capstone/blob/master/VotosMontevideo.xlsx

This screenshot shows the compiled municipalities file:

Municipio	Address	Latitude	Longitude
A	Carlos María Ramírez 862	-34.8657302	-56.2358269
B	Joaquín Requena 1701	-34.8973036	-56.17114
C	L. A. de Herrera 4547	-34.8671485	-56.1714536
X	Brito del Pino 1590	-34.89969	-56.15665
D	Av. Gral. Flores 4694	-34.8455253	-56.1544663
E	Estadio Charrúa	-34.87850575	-56.08935713
F	Av. 8 de Octubre 4700	-34.8587861	-56.1334955
G	Cno. Castro 730	-34.8595754	-56.2134535

This screenshot shows the compiled poll file:

```
dfs.head()
```

	Circuito	SERIE	Municipio	Habilitados	No_Obs	Obs	Emitidos	EnBlanco	Anulados	PorSi	FA	PN	PC	PI	AP	PT	PE	PG	PV	PD	CA	Desde	Hasta	Local	Direccion	Notas
0	1	AAA	B	389	302	8	310	4	3	4	117	87	38	5	2	0	5	4	2	0	31	1	6091.0	Centro SUB OFICIALES NAVALES	WASHINGTON 267 esq. PÉREZ CASTELLANO	NaN
1	2	AAA	B	389	344	4	348	7	8	3	129	81	23	6	4	0	6	2	2	0	73	6092	7223.0	Centro SUB OFICIALES NAVALES	WASHINGTON 267 esq. PÉREZ CASTELLANO	NaN
2	3	AAA	B	389	333	2	335	8	10	1	124	67	35	5	4	1	3	2	5	0	68	7224	8035.0	Centro SUB OFICIALES NAVALES	WASHINGTON 267 esq. PÉREZ CASTELLANO	NaN
3	4	AAA	B	389	317	2	319	4	10	2	133	62	20	5	3	0	3	6	1	0	68	8036	9788.0	MINISTERIO de GANADERÍA AGRICULTURA Y PESCA	CERRITO 318 esq. COLÓN	NaN
4	5	AAA	B	389	338	2	340	1	5	3	164	78	26	4	5	0	2	1	5	1	43	9789	10402.0	MINISTERIO de GANADERÍA AGRICULTURA Y PESCA	CERRITO 318 esq. COLÓN	NaN

Data Dictionary

https://github.com/DSS-TAJorge/Coursera_Capstone/blob/master/DataDictionary.xlsx

Municipalities Data Dictionary screenshot

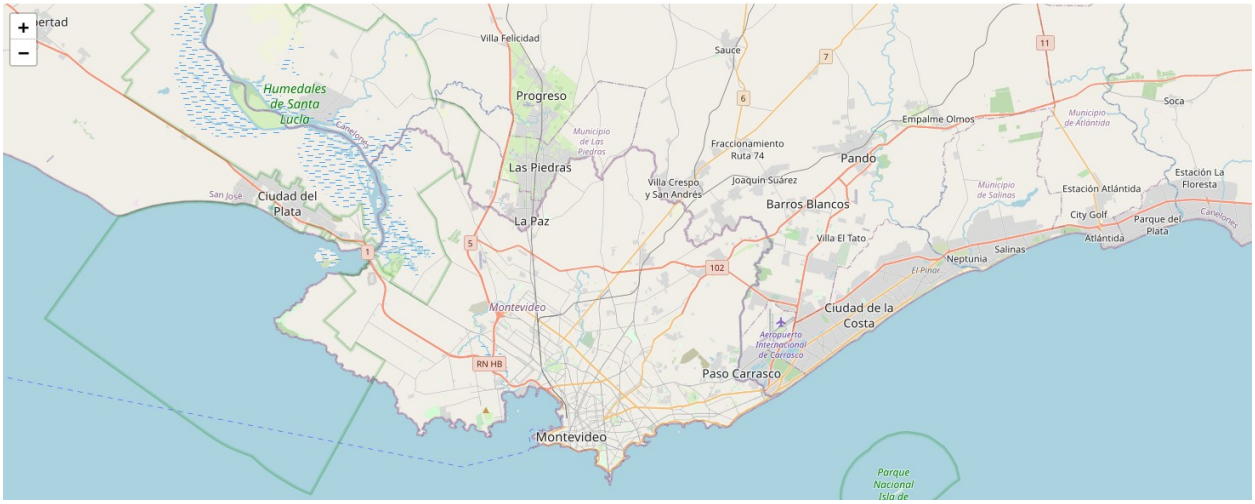
Feature	Meaning	Type	Additional Comments
Municipio	Municipality Code	String	Municipio 'CH' changed to 'X' to keep the same string length
Address	Municipality Center Address	String	Base address of the municipality
Latitude	Latitude of the Address	Float	Extracted used geocode
Longitude	Longitude of the Address	Float	Extracted used geocode

Data Dictionary for the poll file screenshot

Feature	Meaning	Type	Additional comments
Circuito	Poll Station Id	Integer	
SERIE	Document prefix	String	This ia the poll station location for the voters
Municipio	Municipality values A, B, C, D, E, X, F, G or Z	Categorical	X using instead of 'CH' to keep one letter only, Z=special circuit, ignore
Habilitado	Number of voters assigned to the circuit	Integer	
No_Obs	Number of normal votes	Integer	
Obs	Observed votes	Integer	These are votes belonging to other circuits and are not counted until la
Emitidos	Total votes issued	Integer	
EnBlanco	Blank vote	Integer	Vote for no party – issued only because voting is mandatory
Anulados	Null vote	Integer	Incorrect vote that can't be counted
PorSi	Referendum vote	Integer	Vote for Yes – change the Consititution Security Chapter
FA	Votes for party Frente Amplio	Integer	
PN	Votes for party Partido Nacional	Integer	
PC	Votes for party Partido Colorado	Integer	
PI	Votes for party Partido Independiente	Integer	
AP	Votes for party Alianza Popular	Integer	
PT	Votes for party Partido de los Trabajadores	Integer	
PE	Votes for party Partido Ecologista	Integer	
PG	Votes for party Partido de la Gente	Integer	
PV	Votes for party Partido Verde Animalista	Integer	
PD	Votes for party Partido Digital	Integer	
CA	Votes for party Cabildo Abierto	Integer	
Desde	From document number	Integer	These front to numbers, along with the series, determine the poll static
Hasta	To Document number	Integer	These front to numbers, along with the series, determine the poll static
Local	Location of the poll station	String	
Direccion	Address of the poll station	String	
Notas	Comments	String	

Categories and map from Foursquare API:

id	name	pluralName	shortName
4d4b7104d754a06370d81259	Arts & Entertainment	Arts & Entertainment	Arts & Entertainment
4d4b7105d754a06372d81259	College & University	Colleges & Universities	College & Education
4d4b7105d754a06373d81259	Event	Events	Event
4d4b7105d754a06374d81259	Food	Food	Food
4d4b7105d754a06376d81259	Nightlife Spot	Nightlife Spots	Nightlife
4d4b7105d754a06377d81259	Outdoors & Recreation	Outdoors & Recreation	Outdoors & Recreation
4d4b7105d754a06375d81259	Professional & Other Places	Professional & Other Places	Professional
4e67e38e036454776db1fb3a	Residence	Residences	Residence
4d4b7105d754a06378d81259	Shop & Service	Shops & Services	Shops
4d4b7105d754a06379d81259	Travel & Transport	Travel & Transport	Travel



METHODOLOGY

Exploration

Technical Solution Approach

Given the type of insight this study should yield and considering the nature of the available data, in the sense, that we don't have a way to obtain insight via a deterministic approach then **Machine Learning** is the option to take. Within Machine Learning we opted for an **Unsupervised** approach due to the fact that we don't have any previous knowledge or historic data for a regression analysis or to train supervised model.
In terms of algorithms, as we were essentially looking for **Clustering** we've opted for **K-means** because it is the better known clustering method due to its effectiveness and easy of implementation. This proved effective for this study and there was no need to attempt other methodologies.

RESULTS

The output of this study has given us two main assets for discussion. The clustering by venue and service type and the relation between these and the citizen satisfaction in terms of votes

DISCUSSION

It should be noted that although this study has given the results required to be used as a base to find ways of improving the city, we suggest further studies with more data about the municipalities, for example transport and road facilities, neighbourhood economy indicators, etc. In terms of citizen satisfaction it would also be interesting to add actual surveys plus historic data to identify trends.

CONCLUSION

This study has clearly shown the kind of services and venues available in the city separated in three main categories which also are geographically separated. Citizen satisfaction is also related to type of area.
With this study the Government of Montevideo should be able to look into ways of improving the different areas and possible conduct further studies to go deeper in detail of deciding what venues, transport or changes will make a positive impact.