
Test your Business Viability

B. Díaz

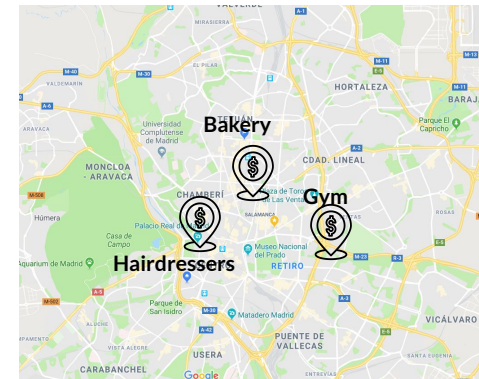
27 de septiembre, 2019

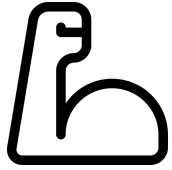


You have a great idea for a business in Madrid. But before you invest your time, capital and resources, you need to figure out whether it's a viable business venture.

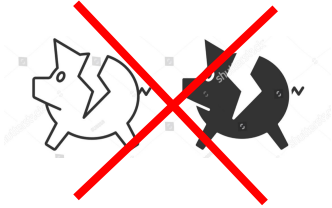
Via analytics, we will try to answer the following questions:

- **Descriptive questions:**
 - understand the impact of population in a business, weekly/weekends
 - Typify zones (postal codes) and representative variables
 - Identify a business influence zone
- **Predictions:**
 - Probability of a businesses to remain opened the year after the opening
 - Type of Business with the highest probability to remain opened the year after in a zone





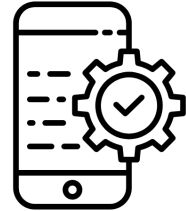
Motivation



In my neighborhood there are locals that have housed a different business every year, sometimes even more than two per year. I have always wondered whether people, mainly self-employees with limited resources, make a proper business viability study prior to invest in their ideas or passions.

For me it is a pity the effort and hopes that many people squander in a business that lasts less than 3 months.

Make use of Madrid public data to help entrepreneurs in Madrid to decide whether to start a new business or not, which one and where





Some statistics

95% of the business are microSMEs (self employees and less than 9 employees)

22,4% of the businesses are retailers with a 96% of micropymes penetration

Retailers represent the biggest businesses destruction (8 k business less YoY)



Main datasets

Censo de locales

- Portal de datos de madrid.es:
- Datos desde 2014 hasta 2018
- Formato csv (fichero/año)
- Id. Local
- Distrito
- Actividad
- Estado
- >5 Millones registros/año

Catastro

- Portal de datos de madrid.es:
- Datos desde 2013 hasta 2017
- Formato csv (fichero/año)
- Distrito
- Tipo inmuebles
- Número inmuebles
- Valor catastral
- 24k registros/año

Padrón

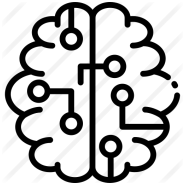
- Portal de datos de madrid.es (banco de datos)
- Datos desde 2014 hasta 2018
- Formato excel/csv
- Población x distrito, barrio, edad, género, nacionalidad, educación
- Gasto anual a nivel Ciudad Madrid x edad, género, educación

Turismo

- Madrid.es y Inside Airbnb
- Airbnb Listings (2019)
- Airbnb Reviews (since 2011)
- Oficinas de turismo
- Tiendas con perfil turístico (xml)
- Puntos de interés turístico (xml)

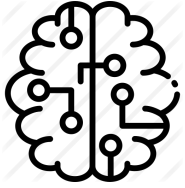
Tráfico / zona

- TBC (NDA)
- Estimación de movimientos de personas/zona/día de la semana
- Anonimizada/agregada

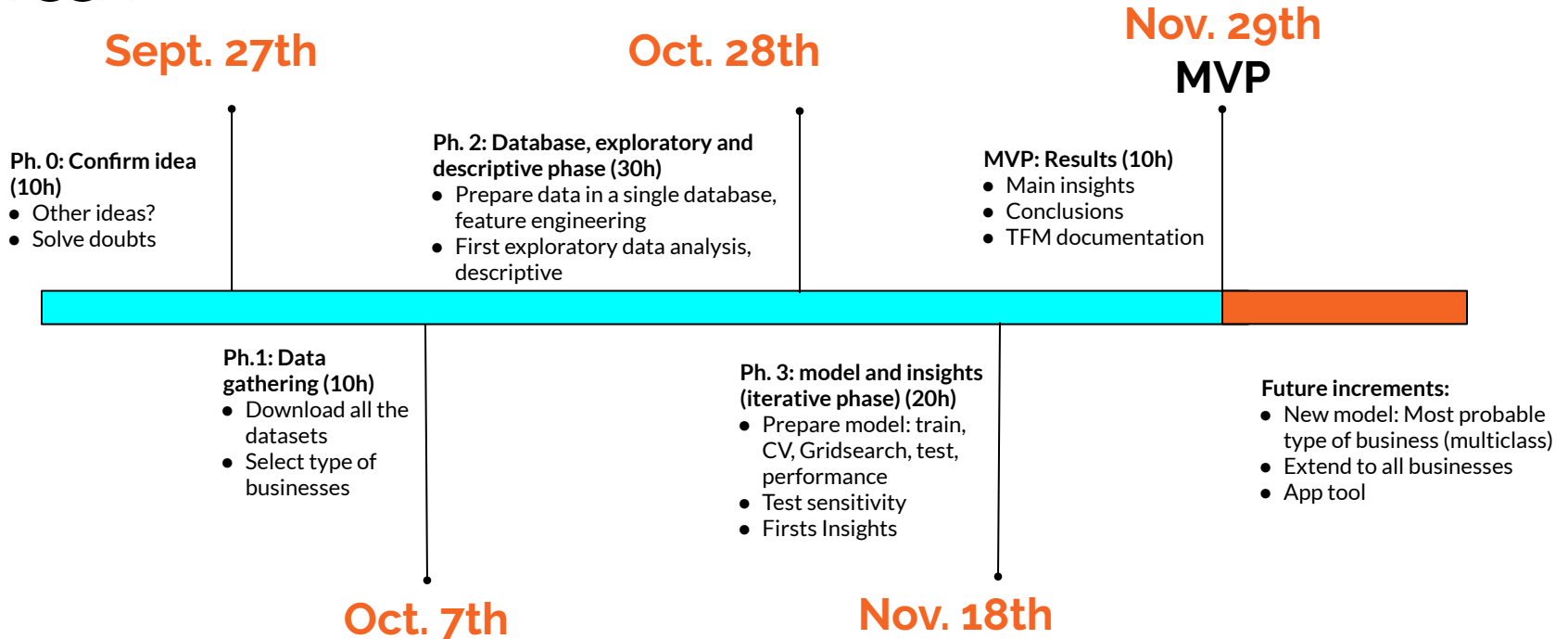


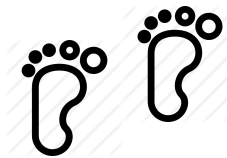
Modeling Methodology: MPV

- **Probability of a businesses to be success**
 - **Type of model:** binary classification problem
 - **Target variable:** Success business = business opened in Madrid the year after the opening year
 - **Independent variables:** type of business, variables that typify a zone (traffic, tourism, sociodemographic characteristics), area of influence of a business (distance between equivalent businesses),
 - Decisions to take **after the first exploratory phase:**
 - Type of business to include in the model based on sample size, baseline/fashion to assure stability
 - socio demographic variability yoy
 - **Train different models:** logistic regression, random forest, XGboost
 - **Test & sensitivity**
 - **Metrics:** AUC, precision
- **Most probable type of business to be success in a zone**
 - Predict per type of business/zone
 - Order prob. per type of business/zone




Project Methodology & Main Milestones





Next steps

- **Clarify doubts:**
 - Error: EmptyDataError: No columns to parse from file (fichero padron)
 - Project doubts tutoring methodology
- **Data gathering and type of business selection** (Due date: October 7th)
- Look for the **5Js shop** closest to my neighborhood 

Follow the project in: <https://github.com/bdiazcor/Kshool-TFM-/>

Appendix

Main datasets

Datos.gob.es (nivel estatal)

- Censo de edificios por tamaño de población: 01008c.csv Fecha ¿?
- Censo de edificios e inmuebles/pobalción: 1mun00sc.csv: censo de edificios e inmuebles/pobalción. Fecha ¿?
- Compraventa de viviendas desde 2007 hasta 2018, por comunidad autónoma y provincia. 6155sc.csv. Interesante
- Empleo total y asalariado Resultados por provincias ramas de actividad magnitud y período (Identificador API: /t35/p010/base2010/I0/02002.px): 02002sc.csv. Años 2012-2016. No muy claro
- Nacimientos por comunidades y ciudades autónomas de residencia de la madre y mes. Movimiento Natural de la Población Resultados Provisionales (Identificador API: /t20/e301/provi/I0/01005.px): 01005sc.csv. Sólo 2019

Censo de locales

<https://datos.madrid.es/sites/v/index.jsp?vgnextoid=23160329ff639410VgnVCM2000000c205a0aRCRD&vgnnextchannel=374512b9ace9f310VgnVCM100000171f5a0aRCRD->historico-de-locales-desde-2014>

Viviendas turísticas: Sección: I, División 55

Padron de madrid, edad, sexo, inmigrantes, movimientos

Datos 2018:

<https://www.madrid.es/portales/munimadrid/es/Inicio/El-Ayuntamiento/Estadistica/Ar-eas-de-informacion-estadistica/Demografia-y-poblacion/Cifras-de-poblacion/Padron-M-unicipal-de-Habitantes-explotacion-estadistica-/?vgnnextfmt=default&vgnextoid=e5613f8b73639210VgnVCM1000000b205a0aRCRD&vgnnextchannel=a4eba53620e1a210VgnVCM1000000b205a0aRCRD>

Histórico desde 2014:

<https://datos.madrid.es/sites/v/index.jsp?vgnextoid=5deb7700f69e8410VgnVCM2000000c205a0aRCRD&vgnnextchannel=374512b9ace9f310VgnVCM100000171f5a0aRCRD>

Banco de datos Madrid.es (si no encuentro algo):

<http://www-2.madrid.es/CSE6/control/menuCSE>

- Gasto anual por educación, sexo, hogar, a nivel de Madrid

INE: Proyecciones de población y de hogares (hasta 2068)

http://www.ine.es/dyngs/INEbase/es/categoria.htm?c=Estadistica_P&cid=1254734710984
Proyecciones de población y de hogares

Catastro - datos desde 2013

<https://datos.madrid.es/portal/site/egob/menuitem.c05c1f754a33a9fbc4b2e4b284f1a5a0/?vgnextoid=3ed1854c8410b410VgnVCM2000000c205a0aRCRD&vgnnextchannel=374512b9ace9f310VgnVCM100000171f5a0aRCRD&vgnnextfmt=default>

Turismo

Airbnb:

<http://insideairbnb.com/get-the-data.html>

10 July, 2019 Madrid listings.csv.gz Detailed Listings data for Madrid

10 July, 2019 Madrid calendar.csv.gz Detailed Calendar Data for listings in Madrid

10 July, 2019 Madrid reviews.csv.gz Detailed Review Data for listings in Madrid

10 July, 2019 Madrid listings.csv Summary information and metrics for listings in Madrid (good for visualisations).

10 July, 2019 Madrid reviews.csv Summary Review data and Listing ID (to facilitate time based analytics and visualisations linked to a listing).

N/A Madrid neighbourhoods.csv Neighbourhood list for geo filter. Sourced from city or open source GIS files.

N/A Madrid neighbourhoods.geojson GeoJSON file of neighbourhoods of the city.

Tiendas con perfil turístico en Madrid

<https://datos.madrid.es/sites/v/index.jsp?vgnextoid=86e3a73970504510VgnVCM2000001f4a900aRCRD&vgnnextchannel=374512b9ace9f310VgnVCM100000171f5a0aRCRD>

https://www.esmadrid.com/opendata/tiendas_v1_es.xml

Oficinas de turismo:

<https://datos.madrid.es/sites/v/index.jsp?vgnextoid=7240e6bfbfe5410VgnVCM1000000b205a0aRCRD&vgnnextchannel=374512b9ace9f310VgnVCM100000171f5a0aRCRD>

Detail Data bank (for municipal register)

No seguro | www-2.madrid.es/CSE6/control/menuCSE

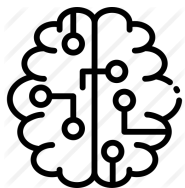
★ Bookmarks  Comenzar a usar Fir...  Importado desde Fi...  Twitter / Inicio  Exercises | Lab: Vari...  Cursos



Está en: [Banco de datos](#)

Áreas de Información Estadística

- ☒ A. Consumo, Precios y Condiciones de Vida
- ☒ B. Cultura, Ocio y Deporte
- ☒ C. Demografía y Población
 - ☒ Población mensual por distritos, barrios y secciones
 - ☒ Padrón municipal
 - ☒ Características de la Población a 1 de enero
 - ☒ Indicadores de la estructura demográfica
 - ☒ Edad
 - ☒ Nacionalidad
 - ☒ Lugar de nacimiento
 - ☒ Hogares
 - ☒ Nivel de estudio
 - ☒ Dinámica Geográfica
 - ☒ Indicadores de la dinámica demográfica
 - ☒ Altas por nacimiento
 - ☒ Bajas por defunción
 - ☒ Inmigración y emigración
 - ☒ Cambios de domicilio en la ciudad
 - ☒ Altas por omisión y bajas por caducidad



Modeling Methodology

- Most probable type of business in a zone in a zone

- Strategy 1 (MPV):

- Model 1 including type of business
 - Predict per type of business
 - Order

Prob. of business Type 1 to be success

Prob. of business Type 2 to be success

Prob. of business Type n to be success

Prob.

+

—

- Strategy 2:

- New model: multiclass classification model
 - Target variable: type of business with highest prob to be success

Prob. to be success



type of business with
highest prob.