

Noya Fernandez del Cotero
Data science Bootcamp
Enero 2023

EDA DUBLIN SCHOOLS

Objetivo

El objetivo de este análisis exploratorio es conocer el mercado de colegios de secundaria de la ciudad de Dublin.

El interés por este proyecto se debe a un interés personal ya que tengo 2 hijos de 12 y 13 años y el año que viene nos mudamos a vivir a Dublin.

Los colegios de secundaria imparten 6 años del sistema educativo repartidos en Junior cycle y Senior cycle .

Junior cycle comienza a los 12 años y senior cycle a los 15 .

La idea es poder responder a preguntas como:

¿Son todos los colegios de Dublín católicos?

¿Son todos gratuitos?

¿Cuál es el precio medio de los colegios de pago?

¿Hay colegios mixtos o son todos de chicos o chicas ?

¿Cuál es el tamaño medio de estos colegios ?

¿Qué barrio es el más familiar, donde hay más colegios ?

¿En qué barrio hay más colegios católicos?

Y sobretodo poder obtener un documento de trabajo con la información general de los colegios: email, web, nombre del director y no tener que ir copiando colegio a colegio desde la web, ya que para poder solicitar la admisión hay que escribir una carta de solicitud de admisión a cada colegio.

Zona	Catchment	Titulo	Principal name	Web	Email	Dirección
Dublin-1	North Inner City	Belvedere College S.J.	Gerry Foley	www.belvederecollege.ie	dreddy@belvederecoll...	6 Gt Denmark Street Dubl...
		Larkin Community College	Thomas Usher	www.larkincommunitycol...	info@larkincommunitycol...	Champions Avenue, Cath...
		Mount Carmel Secondary ..	Mr Gerard Cullen	mountcarmelsecondarysc...	info@mountcarmel.ie	Kings Inn Street, Dublin ...
		O'connell School	Liam Newell	oconnellschool.ie	secretary@oconnellschoo...	North Richmond Street D...
Dublin-2	0	Institute Of Education	Anne Marie Griffin	www.ioe.ie	info@instituteofeducatio...	79/85 Lower Leeson Stre...
	South Inner City	C.B.S. Westland Row	Kate Byrne	www.cbswestlandrow.ie	info@cbswestlandrow.ie	Westland Row Dublin 2 ...
		Catholic University School	Clive Martin	www.cus.ie	office@cus.ie	89 Lower Leeson Street D...
		Loreto College	Jacqueline Dempsey	www.loretothegreen.ie	info@loretothegreen.ie	53 St Stephens Green Du...
Dublin-3	Edenmore-Raheny-Clonta..	Holy Faith Secondary Sch...	Ms Deirdre Gogarty	www.holyfaithclontarf.co...	admin@hollyfaithclontarf...	1 Belgrove Road Clontarf ...
	Finglas	Marino College	Mary McAteer	www.marinocollege.ie	info@marinocollege.ie	14-20 Marino Mart Fairvi...
	Non-Catholic Schools, Dub..	Mount Temple Comprehe...	Liam Wegimont	www.mounttemple.ie	office@md	admin@hollyfaithclontarf.com
	North Inner City	St Josephs C.B.S.	Sean Stack	www.stjosephsfairview.ie	info@stjos	sw Du...
Dublin-4	0	Sandymount Park Educat...	Paula Mulhall	www.sandymountparket...	office@sandymountparke...	The Sandymount Building...
	Sandymount-Ballsbridge...	Marian College	Ms Oonagh McCaul	www.mariancollege.ie	pmeany@mariancollege.ie	Ballsbridge Dublin 4 ,Dub...
		Muckross Park College	Anne Mee	www.muckrossparkcoll...	principal@muckrossparkc...	Donnybrook Dublin 4 ,Du...
		St Conleths College	Donal E O Dulaing	www.stconleths.ie	admin@stconleths.ie	28 Clyde Road Ballsbridg...
		St Michaels College	Mr. Tim Kelleher	www.stmichaelscollege.c...	admin@stmichaelscoll...	Ailesbury Road Dublin 4 ...
		The Teresian School	Deirdre Hickey	www.teresian.ie	secretary@teresian.ie	12 Stillorgan Road Donny...
	South Inner City	Ringsend College	Paul Ryder	www.ringsendcollege.ie	info@ringsend.cdetsb.ie	Cambridge Road Ringsen...

Tras buscar información en la página oficial del gobierno de Irlanda y demás recursos encontré una página donde podía encontrar información y “*en principio*” podía hacer scrapping para obtener estos datos .



Este proyecto se divide en:

1. Scraping de la página "<https://www schooldays.ie>" para obtener un csv con los datos en bruto.
2. Limpieza de los datos
3. Visualización de los datos
4. Presentación del proyecto.

Para el desarrollo de este proyecto la distribución del tiempo ha sido 👍

- 15% Búsqueda del tema a analizar
- 30% web scrapping
- 45% limpieza de datos
- 10% visualización

En cada apartado voy a describir las técnicas que he llevado a cabo para poder ir avanzando y comentando *aprendizajes* y *dificultades* .

1. Scraping

La pagina <https://www schooldays.ie> "Ireland trusted parenting and education in Ireland" es una página de información y comercialización de servicios relacionados con la educación en Irlanda.

Una de las landing agrupa todos los colegios de secundaria por zona .
Vamos a scrapear todos loscolegios de secundaria de la ciudad de Dublin
<https://www schooldays.ie/articles/secondary-Schools-in-Ireland-by-County>;
y dentro podemos acceder a cada zona dentro de la ciudad de Dublin

Ver ejemplo de ruta

1. Elegir mejor la página a scrapear ya que la estructura html de las landing está muy mal hecha , sin etiquetas y en ocasiones no repite el mismo patrón o faltan datos.

```
<b>Post Primary</b>  

<b> School </b>  

"&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;"  

<b> Roll number: </b>  

"60520P -"  

<br>  

<b>:</b>  

<a href="mailto:dreddy@belvederecollege.ie">  

dreddy.belvederecollege.iec/a>  

<br>  

<b>:</b>  

<a href="http://www.belvederecollege.ie" target="_new">  

www.belvederecollege.iec/a == $0  

    br class>  

<b>Principal:</b>  

    "Jerry Enley"
```

Output : Obtenemos un csv *ColegiosDublin.csv* muy sucio....

2. Limpieza de datos

Ver notebook *limpiandoEdaNoya.ipYnb*

El objetivo de este notebook es tratar la información contenida en 10 columnas de strings , de las que me quiero quedar con 2 variables numéricas .

Metodología :

Para abordar esta parte vamos a ir tratando columna a columna , no puedo estandarizar funciones para limpiar ya que ningún string tiene la misma estructura .

Dificultades y aprendizajes :

1. Las columnas :Teléfono,Principal name ,Información general , Ethos y Catchment serán columnas categóricas ,donde la mayor dificultad fue ver que faltaban datos y cuando rompíamos el string , asumiendo que cada bucle asignaba en la posición el contenido que le correspondía ...no funcionaba → lo solucionamos con try except y en varias ocasiones nos dimos cuenta al visualizar y tuvimos que volver a tratar la información con máscaras .

2. Enrolment :

Partíamos de una serie con celdas :

`Enrolment:_Boys:_1003 (2022/23]`

`Enrolment:_Boys:_204 Girls:_203 (2022/23`

`Enrolment:_Girls:_399 (2022/23)`

Usando máscaras consigo , romper la serie original en 4 mini dataframes, chicas , chicos , chicas mixtos y chicos mixtos que luego uno con el método left join vuelvo a unir en dos nuevas columnas con los datos de enrolment de chicas y chicos por separado

3. "", encontrar un string vacío , ha sido muchas veces trampa !! por que paraba el bucle sobre en que iterar , al final lo hemos resuelto o bien reemplazando con "0" o bien convirtiendolos en Nan y luego sustituyendo los NaN por ceros .

En esta etapa del proyecto ha sido vital contar con stackover flow y con la ayuda de Ta y compañeros y por supuesto , revisando los documentos de clase.

Output :df2.limpio.xl

3. Visualización

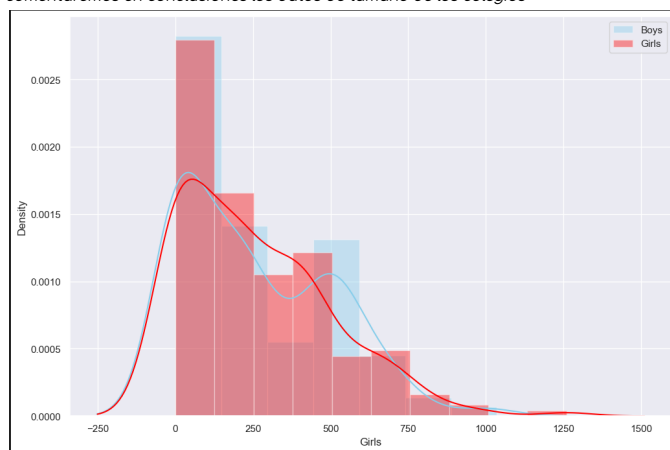
Ver notebook *limpiandoEdaNoya.ipynb*

Analizando variables numericas y categoricas ejemplos:

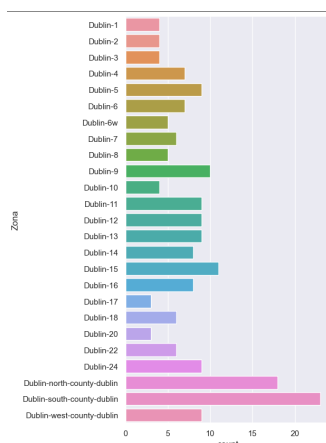
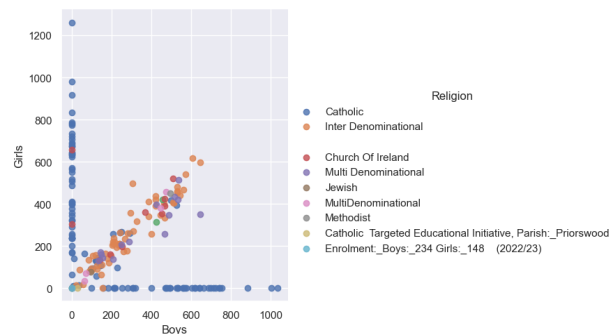
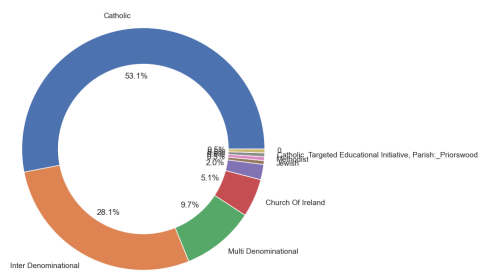
Hemos usado la libreria seaborn para sacar gráficas estáticas que representen los datos .

Con los histogramas y la linea de distribución de densidad vemos pocas diferencias en el tamaño de los colegios de chicos y chicas ,siendo la mayor parte de los colegios de menos de 250 alumnos .

Representación del número de alumnos chicos y chicas
comentaremos en conclusiones los datos de tamaño de los colegios

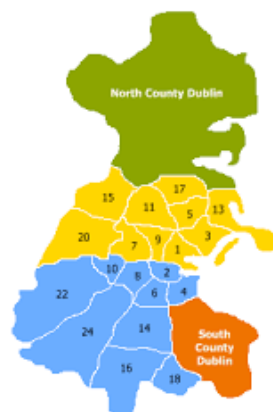


En este gráfico de círculo vemos los tipos de colegios según la religión .
y un scatter plot de religion vs género, donde apreciamos que los colegios mixtos son principalmente interdenominacional.



más

más

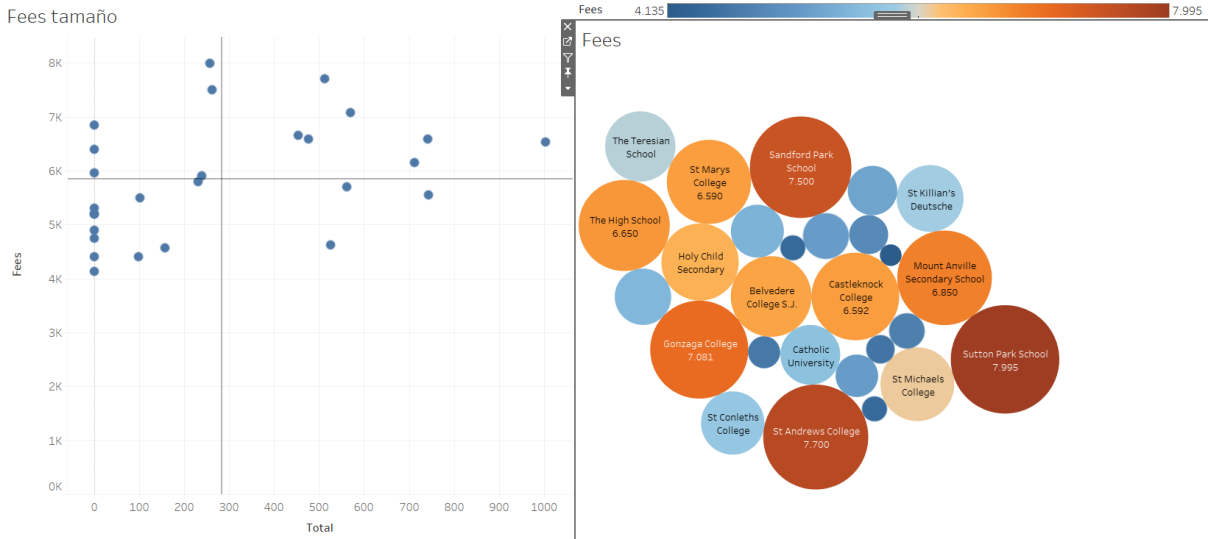


Zonas con más colegios → parece que north county y south county son más familiares ya que tiene muchos colegios. Zonas con más colegios → parece que north county y south county son más familiares ya que tiene muchos colegios.

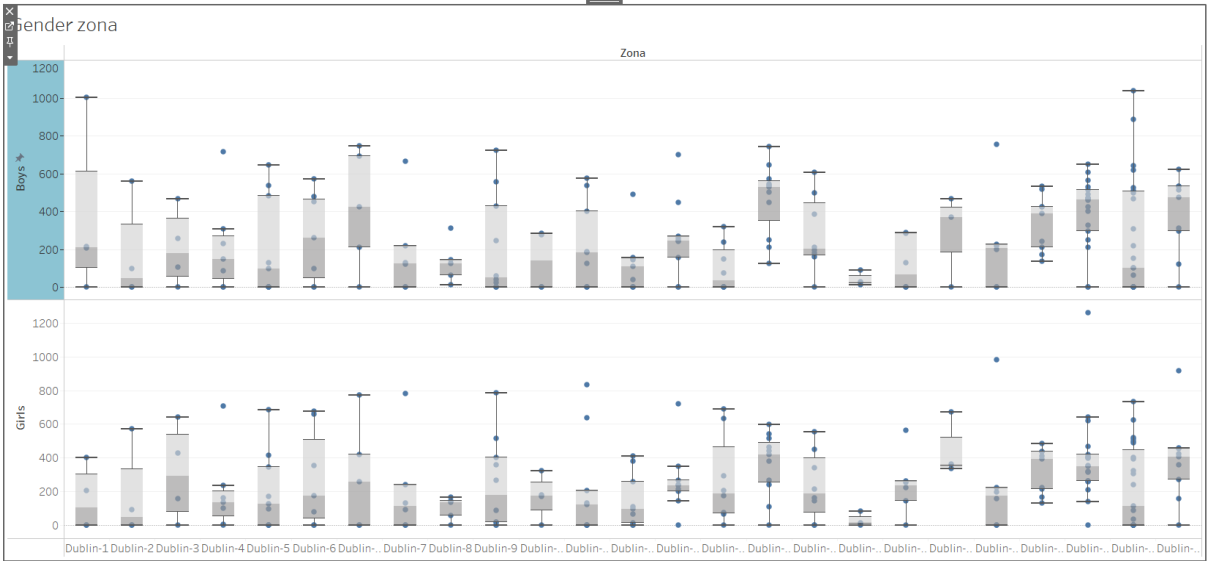
Ver proyecto Tableau

¿Son todos los colegios en Dublin privados? ¿Cual es el colegio más caro?
Solo con los 28 colegios de pago, representamos tamaño del colegio vs precio. Y en burbujas las tarifas con el nombre del colegio.

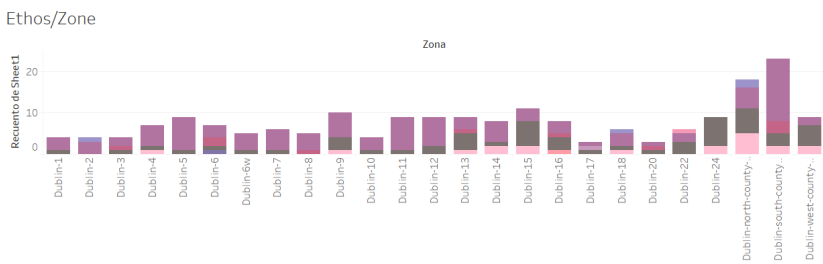
El precio medio de los colegios privados es de 5.446€ y el más caro 7.995€ es Sutton Park school y el más económico Loreto High school Beaufort con 4.135€



¿Cual es el tamaño promedio de los colegios? Hay diferencias entre los colegios de chicas o chicos?
En la distribución de las zonas vemos los outliers sobretodo en zonas North county en boys con Blackrock college con 1036 alumnos, y Loreto school con 1260, también destaca sobresaliendo sobre la horquilla de su zona en Dublin 13 los colegios Fintahight y Santa Sabina en boysand girls respectivamente.



¿Son todos los colegios de Dublín católicos? ¿En qué barrio hay más colegios católicos?



Conclusiones:

1. Hay 50.833 chicos y 50.527 chicas escolarizados en colegios de secundaria en Dublin.
2. En 196 colegios que se reparten en 27 zonas con 48 Catchment areas diferentes .
3. Las zonas con menos colegios son Dublin 17 y Dublin 20 con solo 3 colegios cada una en contraste con Dublin norte y Dublin Sur que tienen 23 y 18 colegios cada una representando un 14% y 11% de los niños escolarizados en secundaria .
4. No todos los colegios son católicos, de hecho representan el 53% con un 50% de los alumnos matriculados ,seguidos de los Inter denominational (donde conviven católicos , protestantes y metodistas) con un 30% ; los multidenominational son aquellos donde la fe no se enseña en el colegio sino se deja en manos de los padres y suponen un 9.7% y los protestante de Church of Ireland un 5.1%
5. Solo hay 174 alumnos judios matriculados en Stratford college , que es de pago y cuesta 4400€
6. Un 14.7% son colegios de pago , la zona South county Dublin es donde más colegios de pago hay .
7. El precio medio de los colegios de pago es de 5.446€ y el más caro 7.995€ es Sutton Park school y el más económico Loreto High school Beaufort con 4.135€
8. El tamaño de los colegios es de media de 517 alumnos ; siendo Loreto secondary School con 1260 niñas el más grande y Castleknock Community College con 1224 chicos el siguiente .
9. Belcamp college con 27 alumnos matriculados en Dublin 17 es el colegio más pequeño .

Ideas erróneas

“Todos los colegios de Dublin son católicos”

“ Todos los colegios son de pago”

“Los colegios de pago son solo de chicos o de chicas “ hay 8 colegios mixtos de pago