

# Data Analysis

Data Science Bootcamp

The Bridge









# Exploratory Data Analysis

*¿De qué se compone un EDA?*



## Hipótesis

*Cuestiones de negocio como puntos débiles, oportunidades*



## Estructura

*Cómo son tus datos, de qué tipo son, calidad del dato*



## Limpieza

*Datos duplicados, missings, features correlacionadas.*



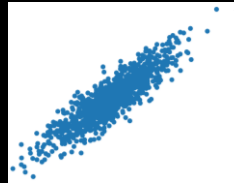
## Outliers

*Técnicas de detección e imputación de outliers*



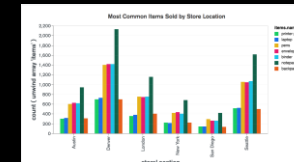
## Feature Engineering

*Transformación de variables y creación de nuevas a partir de las originales*



## Relación entre datos

*Qué variables están correlacionadas*



## Visualización

*Representación de relaciones y conclusiones sacadas de los datos.*

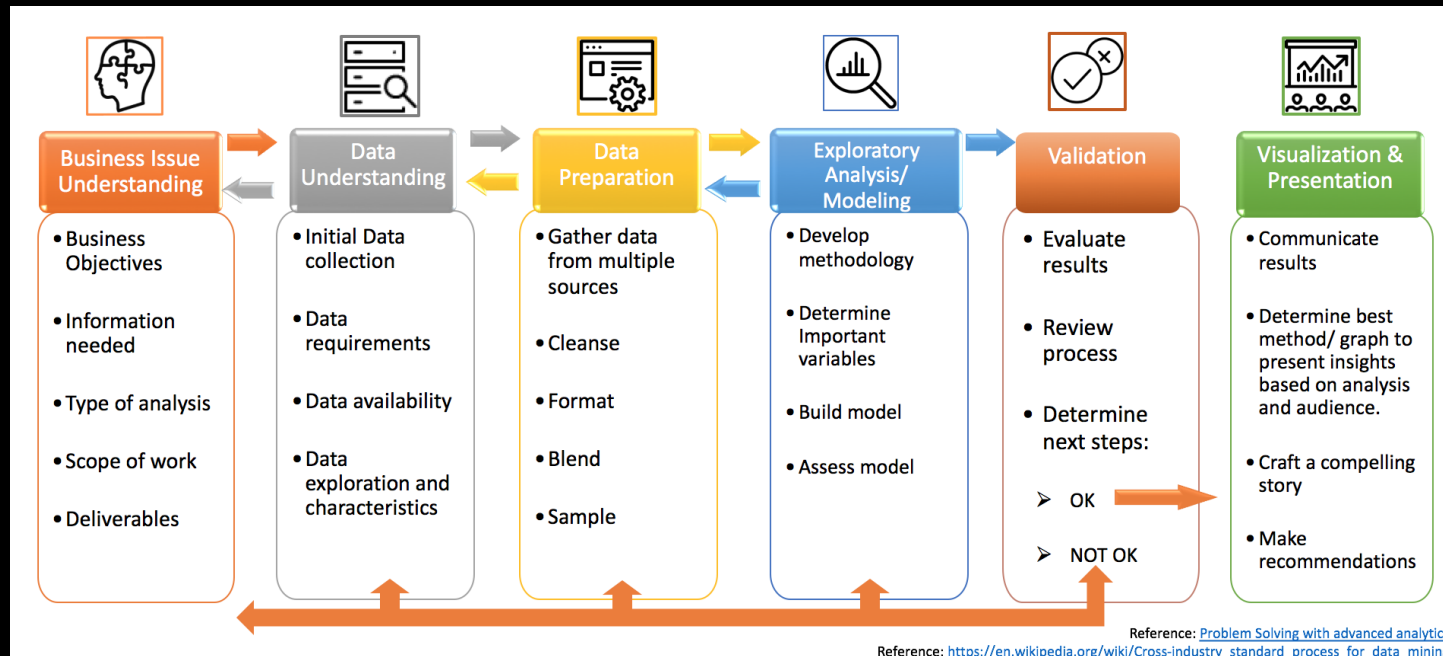
# Proyecto Data Science

*¿En qué momento se hace un EDA?*



# Proyecto Data Science

*¿En qué momento se hace un EDA?*



## ¿Cuándo se realiza un EDA?

*Preparamos los datos para realizar un modelo de machine learning*

*Análisis de datos de la empresa para entender el negocio y los clientes.*

*Análisis de la calidad de los datos*

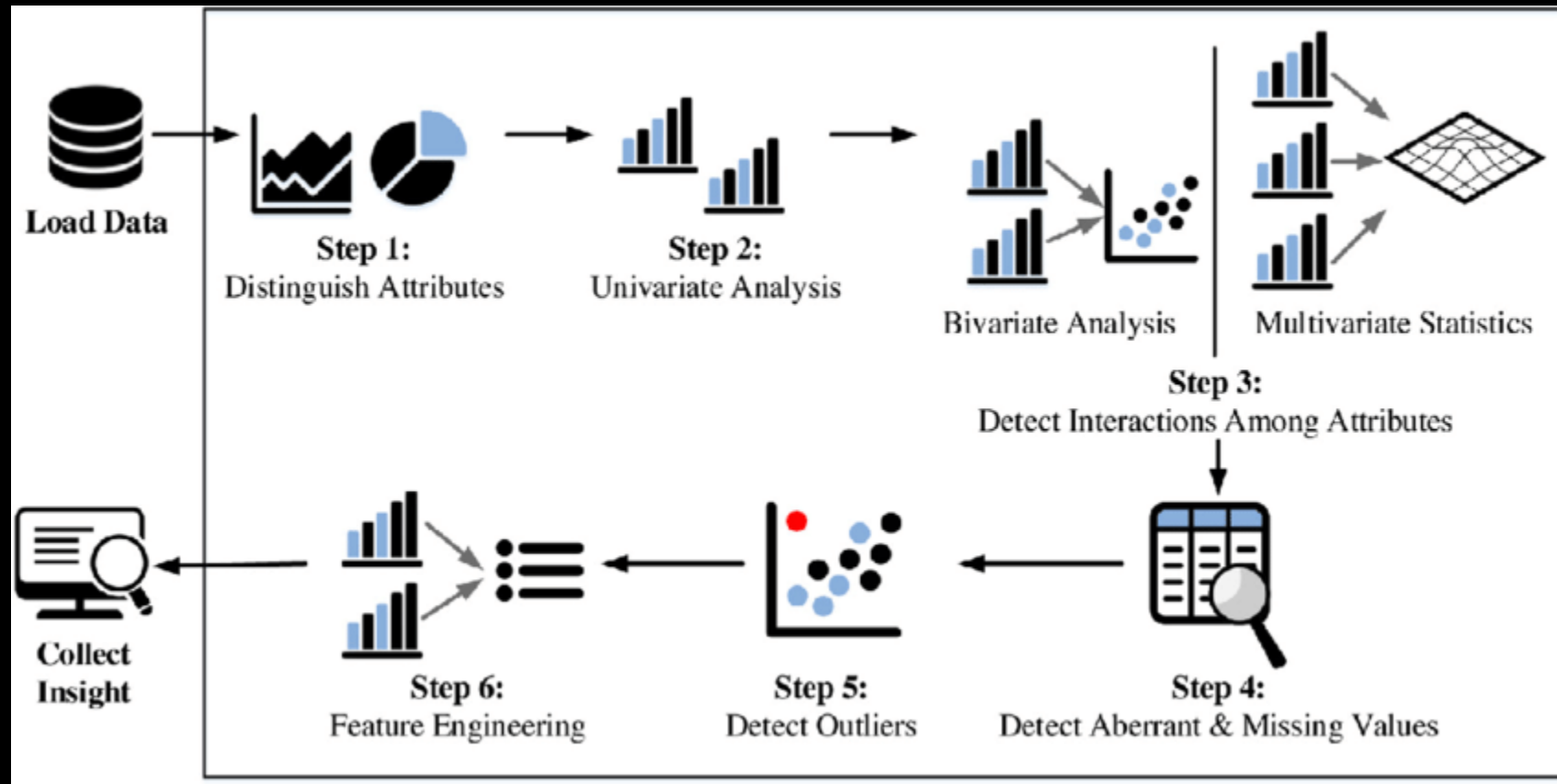
*Reportes*

*¿Ha tenido éxito el nuevo producto lanzado? Análisis de ventas y de campañas de marketing*

*¿Con qué producto estamos teniendo más pérdidas? ¿O más ganancias?  
¿Cómo es el cliente que compra esos productos?*

# Pasos a seguir

*Elementos imprescindibles de un EDA*



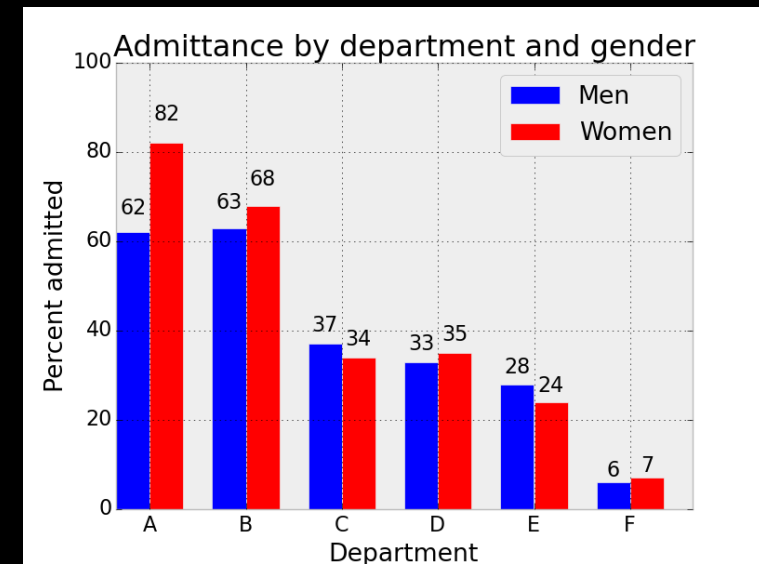
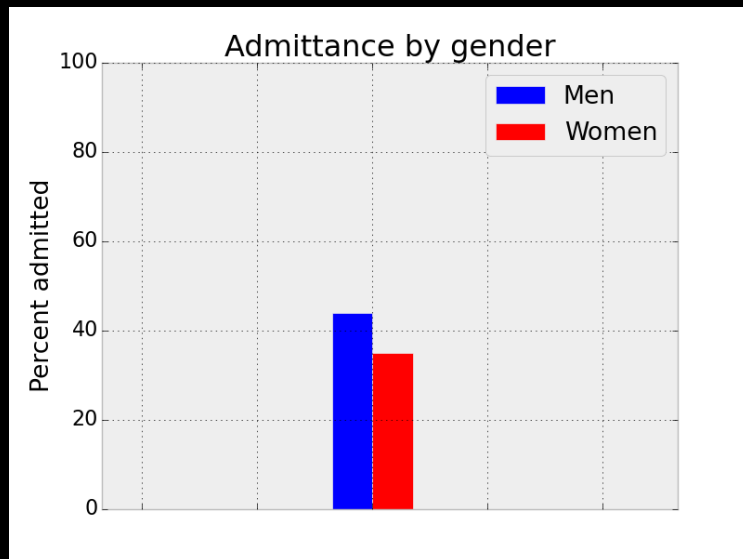
# ¿Discriminación de sexo en Berkeley?

## *Paradoja de Simpson*

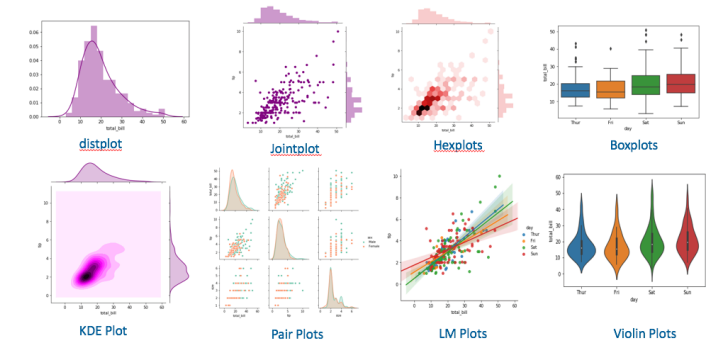
En 1973, la Universidad de California-Berkeley fue demandada por discriminación por género.

En realidad lo que ocurrió fue que al ver las siguientes cifras de admitidos el decano pensó que podrían ser demandados y pidió al profesor de estadística Peter Bickel que analizará la situación.

	Males		Females	
	Applicants	% Admitted	Applicants	% Admitted
A	825	62%	108	82%
B	560	63%	25	68%
C	325	37%	593	34%
D	417	33%	375	35%
E	191	28%	393	24%
F	373	6%	341	7%



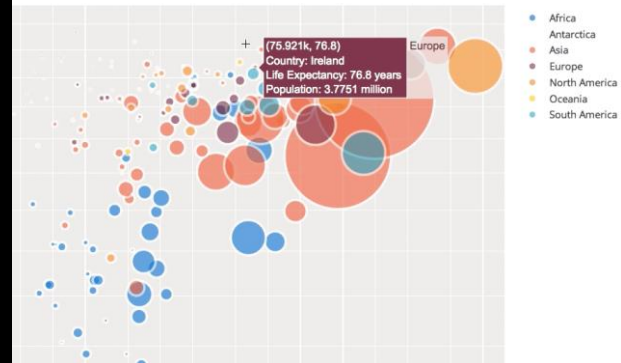
## Seaborn Plots



## Visualizaciones con Python

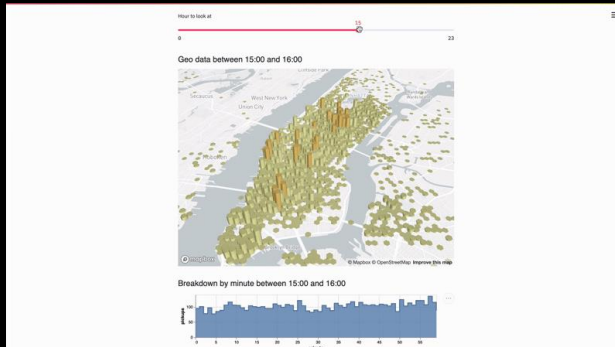
Matplotlib y Seaborn

## Life expectancy vs GNP from MySQL world database (bubble chart)



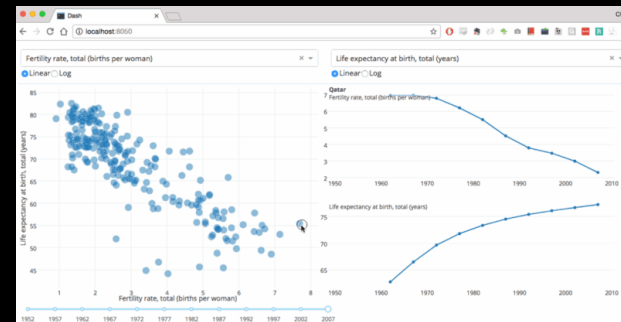
## Informes en HTML

Plotly



## Dashboards sencillos Python

Streamlit



## Dashboards en Python

Dash  
[Dash gallery](#)



## Dashboard Business Intelligence

Power BI



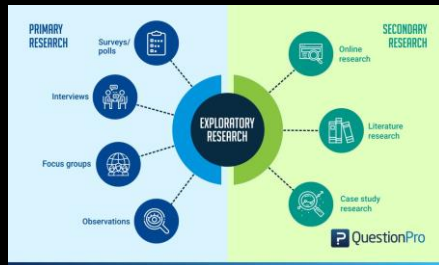
# Presentación de un EDA

*Dependerá de tu audiencia*

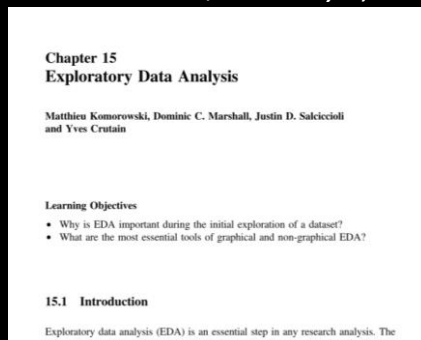


# Presentación de un EDA

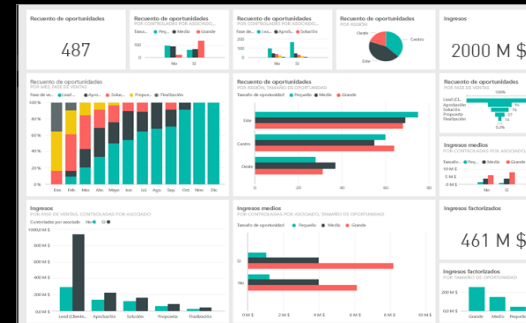
*Dependerá de tu audiencia*



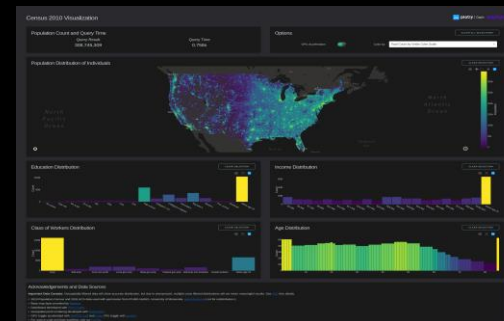
*Presentación (PowerPoint, Prezi, WordPress, MS Sway...)*



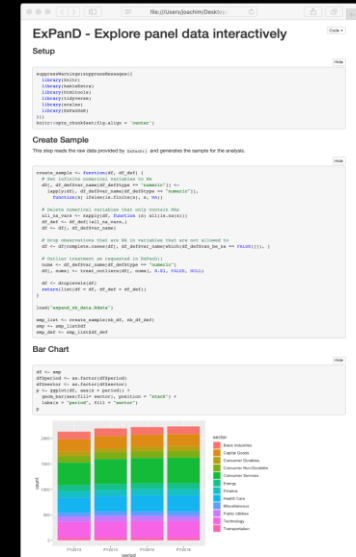
*Informe (Word, pdf)*



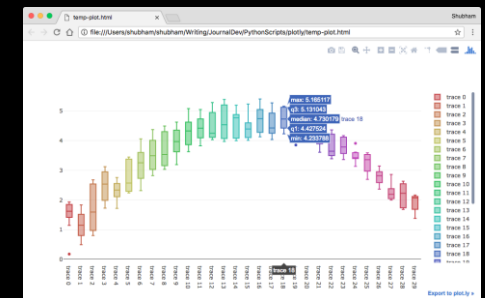
*Dashboard en herramienta de BI*



*Dashboard en Python*



*Notebook técnico*



*HTML con Plotly*

**Presentación de resultados  
Negocio, cliente**

**Productivo  
Muchos usuarios**

**Técnico  
Interno**

# | Bootcamp



seaborn

