

# El arte de la analítica

José Ignacio Treviño  
Luz Eunice Ángeles



# Acerca del curso

Durante esta semana Tec experimentarás con el proceso de análisis de datos que permite resolver preguntas acerca de ellos.

En esta experiencia aprenderás a:

1. Realizar el tratamiento de datos y análisis rápido de un conjunto de datos reales.
2. Obtener información relevante a partir de los datos y presentarla ante el grupo.

# Acerca de los instructores



**Ing. Luz Eunice Angeles**

- Maestría en Admin. Tecnologías de Información
- Correo electrónico: luzeunice@tec.mx
- Mensajería instantánea (Remind ):  
<https://www.remind.com/join/>



**Ing. José Ignacio Treviño**

- Maestría en Investigación de Operaciones
- Correo electrónico: josei.trevino@tec.mx
- Mensajería instantánea (Remind ):  
<https://www.remind.com/join/>

# Contenido del curso

**01**

## Introducción

Unix y Git

**02**

## Herramientas

Entorno de desarrollo en la nube

**03**

## Pandas

Librerías para manejo de datos

**04**

## Estadísticas

Análisis descriptivo

**05**

## Gráficos

Visualización de datos

**06**

## Regresión lineal

Analítica predictiva



02

# Herramientas

desarrollo en la nube

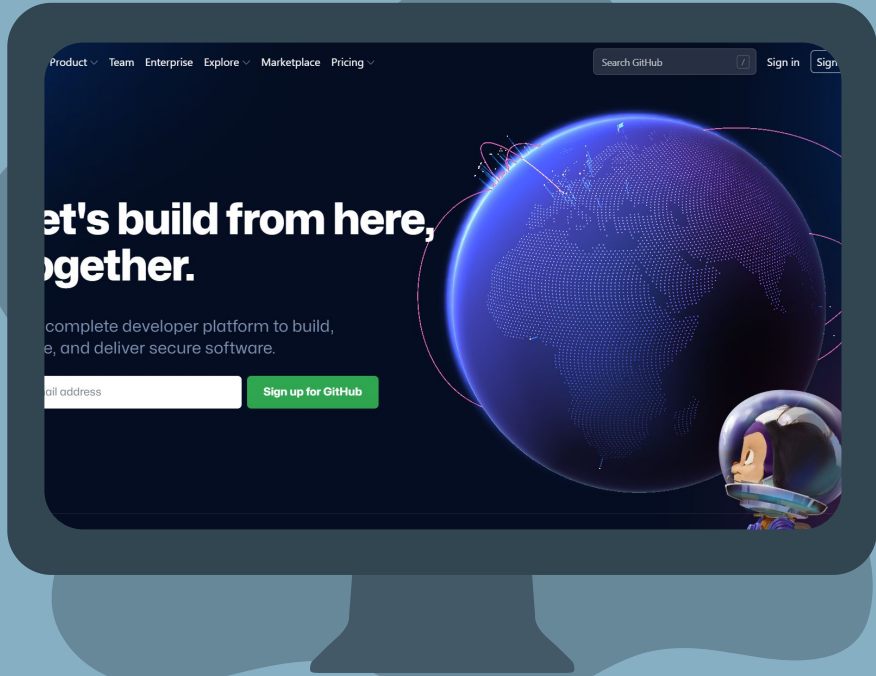
# Google Colaboratory

Es una herramienta que permite ejecutar código de Python en servidores de Google.

Es como Jupyter Notebook que corre en la nube.

[Te damos la bienvenida a Colaboratory - Colaboratory \(google.com\)](#)



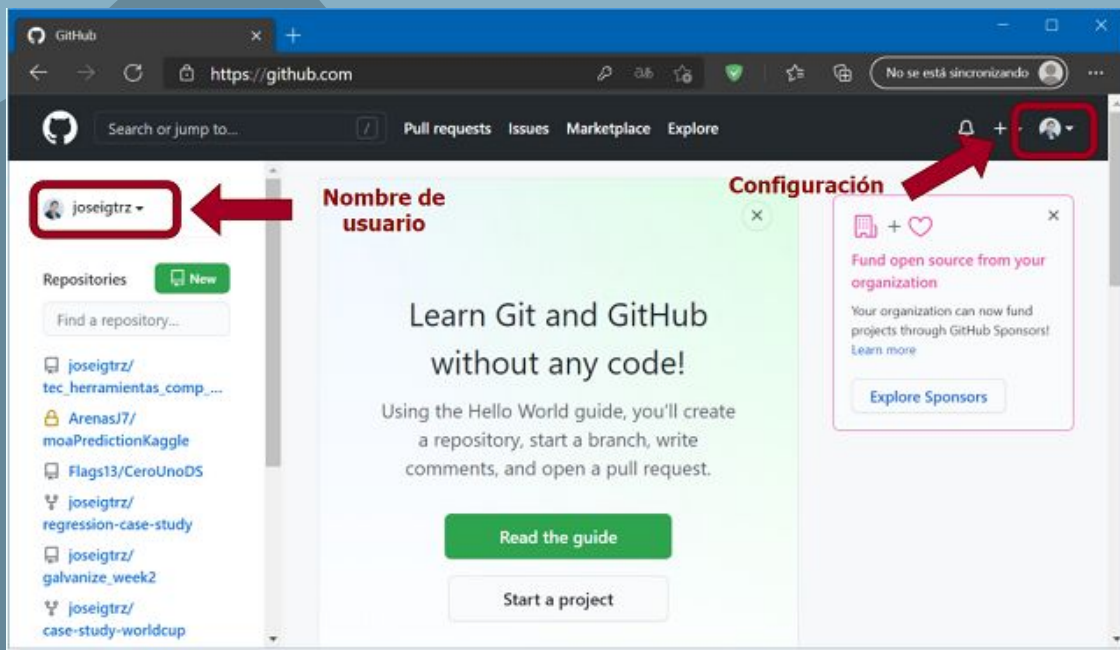


# Github

Es un sitio para alojar proyectos utilizando el sistema de control de versiones Git.

[GitHub](#)

# Conectar Colab con Github



Para poder conectarnos a Github desde Colab, requerimos tres datos:

- Nombre de usuario
- Nombre del repositorio
- Token de acceso



# Notebook de teoría y actividad

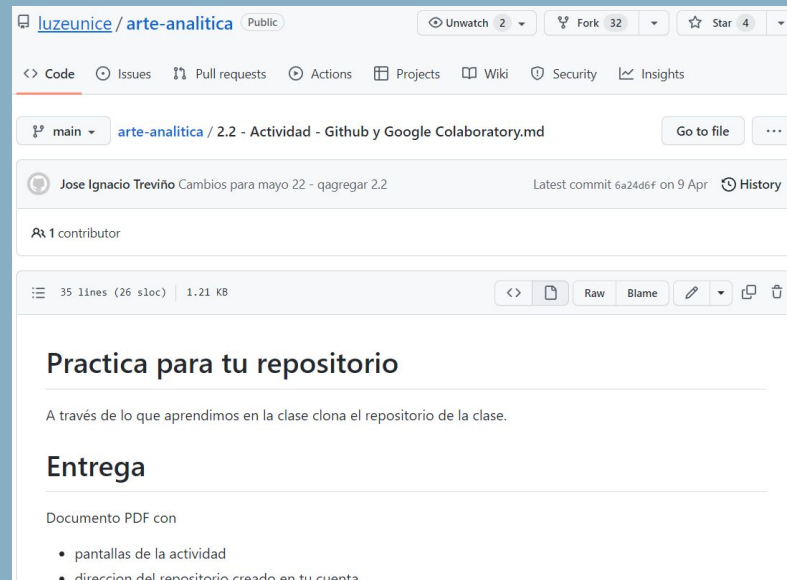


Liga a la libreta de teoría



Liga a la actividad

- Descargar desde Canvas los notebooks y el archivo “credenciales.json”

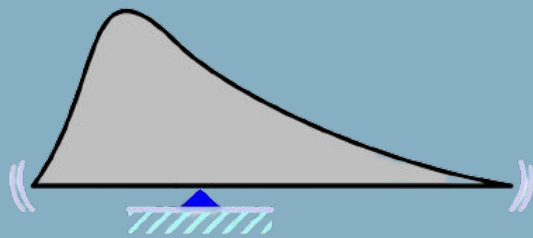




04

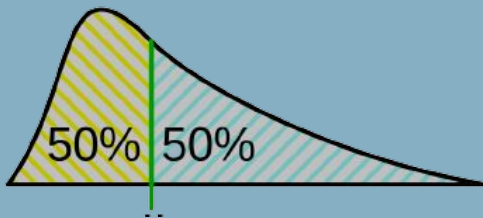
# Estadísticas

# Medidas de tendencia central



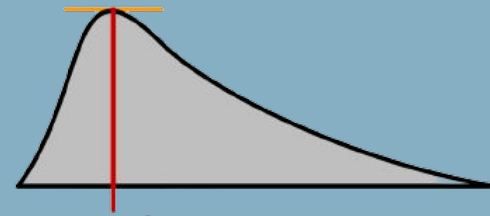
## Media

Es el promedio aritmético



## Mediana

Valor central de datos ordenados.  
2do cuartil / percentil 50



## Moda

Valor más frecuente

# Medidas de dispersión

## Varianza

Promedio de variación con respecto a la media

## Desviación estándar

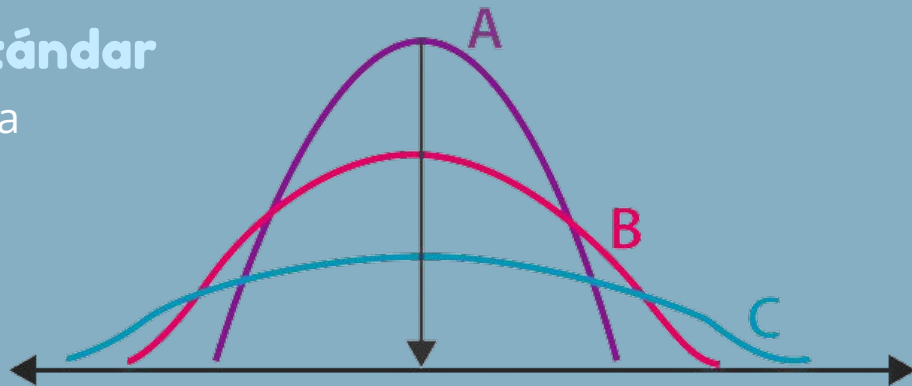
Raíz cuadrada de la varianza

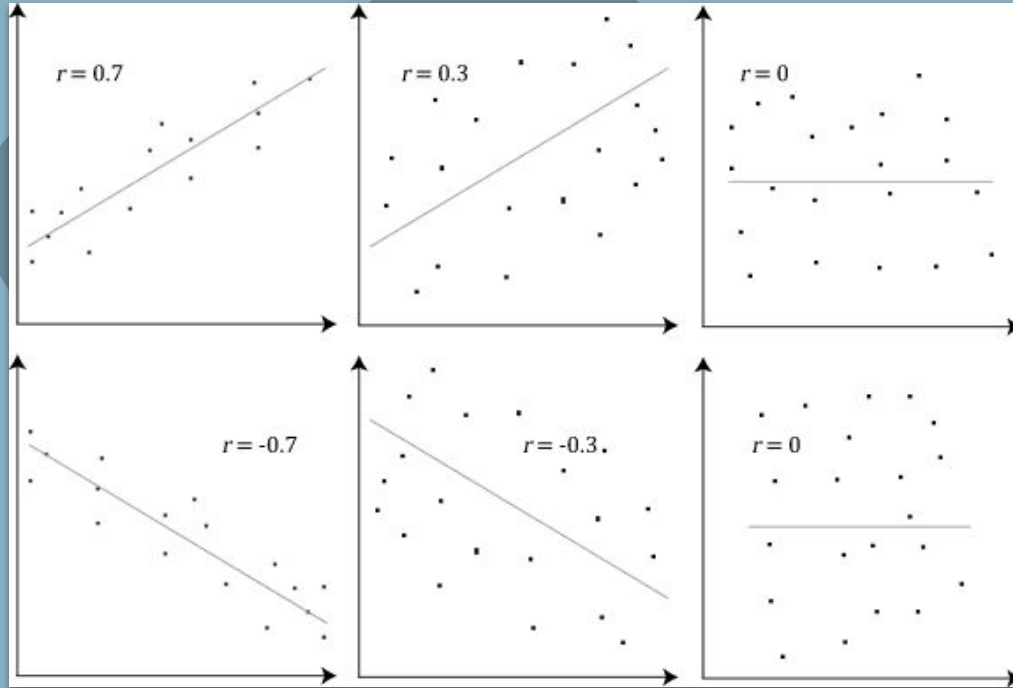
## Rango

Diferencia entre el valor máximo y el mínimo

## Rango intercuartil

Diferencia entre el 3er y 1er cuartil





# Coeficiente de correlación

Es un valor que indica la fuerza y dirección de la relación lineal que hay entre dos variables.

Varía desde -1 hasta 1.

# Estadísticas de datos con Python

Para obtener las estadísticas generales de un Data Frame usamos

- `df.describe()`

Para obtener el promedio de una columna usamos:

- `df['col'].mean()`
- Otras estadísticas que podemos usar son: `sum()`, `std()`, `var()`, `min()`, `max()`, `median()`.

Los coeficientes de correlación se obtienen con:

- `df.corr()`

La moda la podemos obtener contando los diferentes valores de una columna:

- `df['col'].value_counts()`

# Notebook de teoría y actividad



[Liga a la libreta de teoría](#)

luzeunice / arte-analitica Public

<> Code Issues Pull requests Actions Projects Wiki Security Insights

main arte-analitica / 4.1 - Estadísticas de datos.ipynb Go to file

Jose Ignacio Treviño Cambios para mayo 22 Latest commit f4d75ba on 2 Apr History

1 contributor

1 lines (1 sloc) | 13.6 KB

## Estadísticas de datos con Pandas

Semana Tec

Profesores: José Ignacio Treviño / Luz Eunice Angeles

```
In [ ]: # Si trabajamos en Google Colaboratory corremos las siguientes líneas de código
from google.colab import drive
drive.mount('/gdrive')
```

```
In [ ]: # Nos cambiamos a la carpeta donde tengamos el repositorio
%cd /gdrive/MyDrive/SemanaTec/arte-analitica
```



[Liga a la actividad](#)

luzeunice / arte-analitica Public

<> Code Issues Pull requests Actions Projects Wiki Security Insights

main arte-analitica / 4.2 - Actividad - Estadísticas de datos.ipynb Go to file

Jose Ignacio Treviño Cambios para mayo 22 Latest commit f4d75ba on 2 Apr History

1 contributor

1 lines (1 sloc) | 5.23 KB

## Actividad

(Haz doble clic en esta celda y completa los datos correspondientes a tu equipo)

- Integrante 1
- Integrante 2
- Integrante 3

## Entregar

Archivo PDF de la actividad y la liga de la actividad en su repositorio.

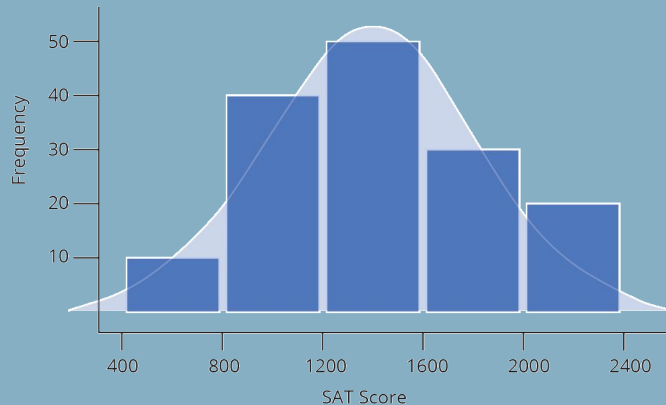
05

# Gráficos



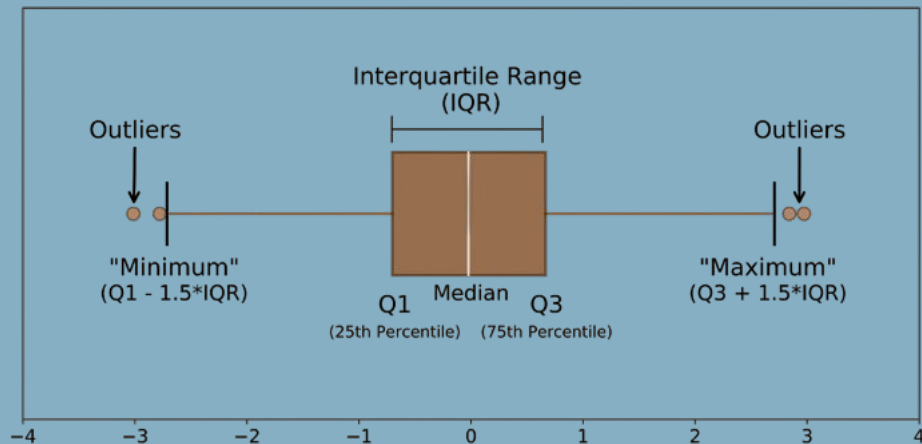


# Gráficos para variables numéricas



**Histograma**

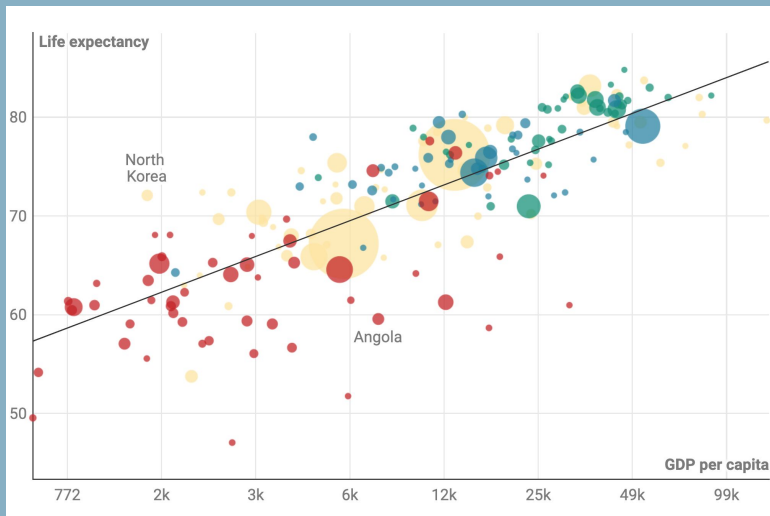
Es una estimación de la densidad de una variable numérica.



**Gráfico de caja**

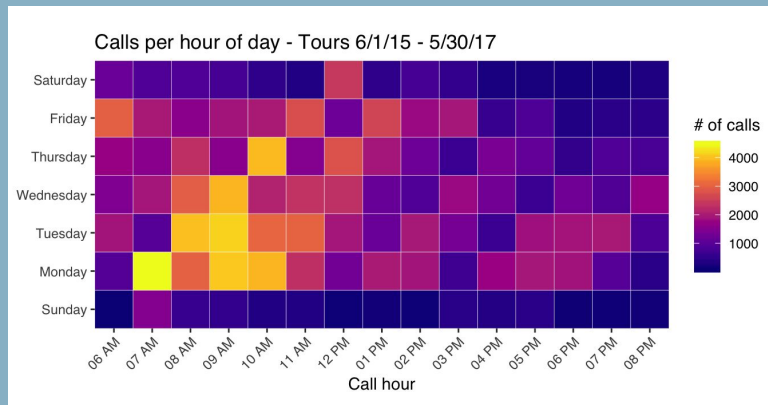
Utiliza los cuartiles para describir una variable numérica.

# Gráficos para variables numéricas



## Dispersión

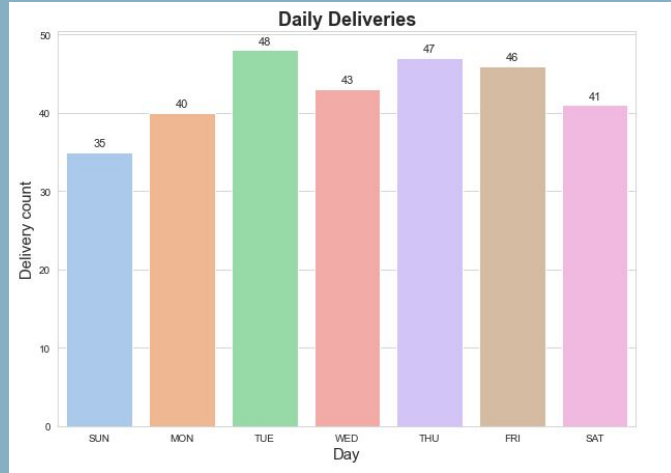
Muestra la relación de dos variables numéricas



## Mapa de calor

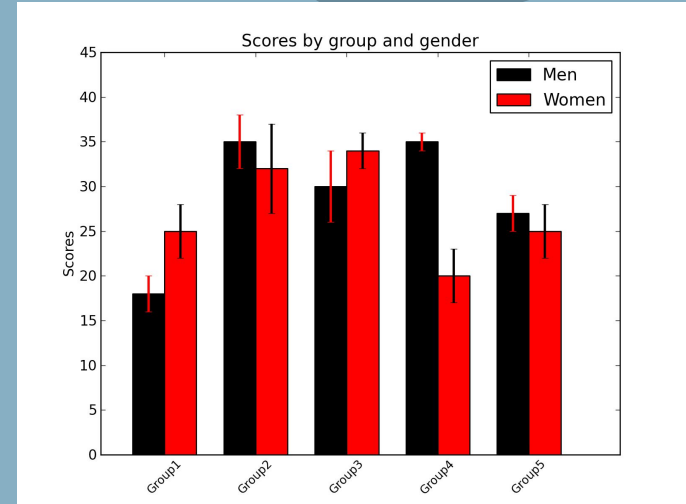
Representación gráfica de una tabla de dos variables

# Gráficos para variables cualitativas



## Gráfico de barras

Para contar las categorías de una variable cualitativa



## Gráfico de barras con estimador

Agrupar una variable numérica por las diferentes categorías de una variable cualitativa

# Visualización de datos con Python

Usaremos la librería Seaborn: `import seaborn as sns`

La estructura general del código para generar un gráfico es la siguiente:

- `sns.<tipo de gráfico>(data = <df>, x = <col x>, y = <col y>, hue=<col color>, style = <col estilo>, size = <col tamaño>, ... )`

Los tipos de gráfico que podemos hacer son:

- `histplot` para histogramas
- `boxplot` para gráficos de caja
- `scatterplot` para gráficos de dispersión
- `heatmap` para mapas de calor
- `countplot` para gráficos de barra de una variable.
- `barplot` para gráficos de barra con dos variables (1 cualitativa y 1 numérica)

# Notebook de teoría y actividad



[Liga a la libreta de teoría](#)



[Liga a la actividad](#)

luzeunice / arte-analitica Public

<> Code Issues Pull requests Actions Projects Wiki Security Insights

main arte-analitica / 5.1 - Visualizacion de datos.ipynb Go to file

Jose Ignacio Treviño Cambios para mayo 22 Latest commit fdd75ba on 2 Apr History

1 contributor

1 lines (1 sloc) 43.4 KB

## Visualización de datos

Semana Tec

Profesores: José Ignacio Treviño / Luz Eunice Angeles

### Librerías de visualización de datos

- **Matplotlib** es la librería más utilizada para crear gráficos en Python.
  - Está basada en la graficación de Matlab.
  - Permite agregar elementos sobre el gráfico, modificar títulos y leyendas, cambiar colores, entre muchas otras

luzeunice / arte-analitica Public

<> Code Issues Pull requests Actions Projects Wiki Security Insights

main arte-analitica / 5.2 - Actividad - Visualizacion de datos.ipynb Go to file

Jose Ignacio Treviño Cambios para mayo 22 Latest commit fdd75ba on 2 Apr History

1 contributor

1 lines (1 sloc) 5.48 KB

## Actividad

(Haz doble clic en esta celda y completa los datos correspondientes a tu equipo)

- Integrante 1
- Integrante 2
- Integrante 3

## Entregar

Archivo PDF de la actividad y la liga de la actividad en su repositorio.

# ¡Gracias!

## Terminamos la Semana Tec

Recomendaciones para finalizar

- La fecha límite para subir actividades es el 16 de septiembre a las 23:59.
- La evidencia es un trabajo en equipo, pero se tiene que subir **individualmente** a eLumen.
- Favor de contestar la coevaluación y la ECOA.

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon** and infographics & images by **Freepik**.

Please keep this slide for attribution