

Acceso a Drive de Master:

<https://drive.google.com/drive/folders/1Djwz81vEHw3lwtErzBYiZml9yFGI2Q-E?usp=sharing>

Modulo 1

220531 Clase 1

Senseis David Ramos, Yeifer Rodríguez

Presentación.-

https://docs.google.com/presentation/d/1wta_zmA0gnpbY8CFF_8duvKcnns49DDrX2eTUZiGqEs/edit#slide=id.g110b46be83c_3_618

Libros interesantes:

Grit de Angela Duckworth ejemplo de westpoint,

Atomic Habits de James Clear

Introducción a Python: https://colab.research.google.com/drive/1r9UjmgT9R5iUd_sa6C70ZjIXN1aJVSc0

INTRO A LA CIENCIA DE DATOS

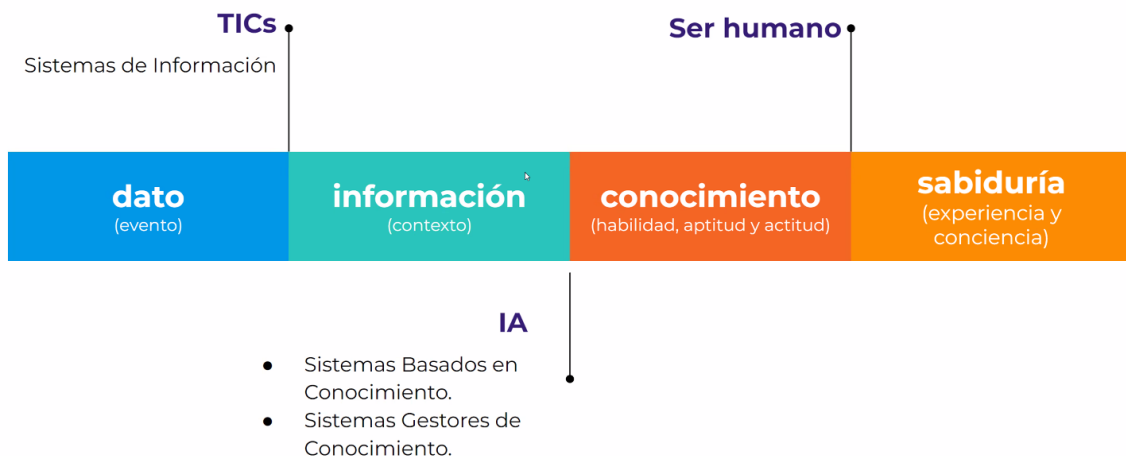
<https://docs.google.com/presentation/d/18XXWb6bI4zb-UZhEsZsLKosIY6g62WbulxBPZlboa3E/edit#slide=id.p1>

Primer Tarea.- Mujeres en programación

ORIGENES: Jhon W. Tukey 1962 Artículo "The future of Data Analysis", 1977 "Exploratory Data Analysis", PRUEBA DE Tukey para el análisis de la varianza(ANOVA)

Dato => Información => conocimiento => sabiduría

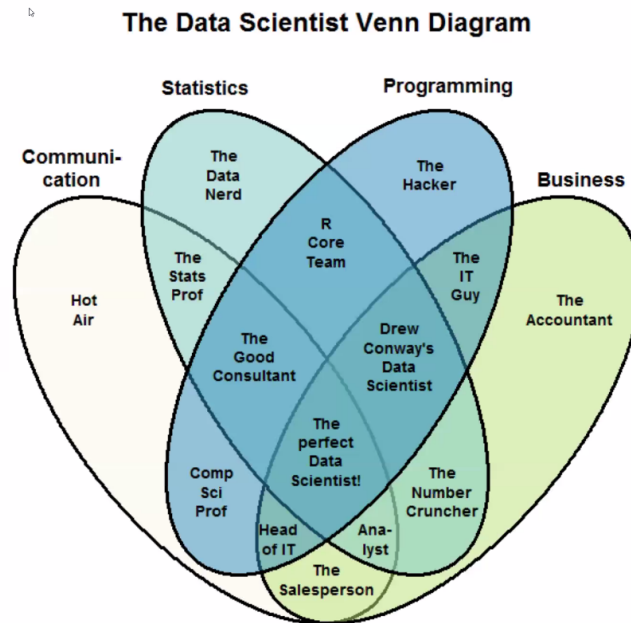
Camino hacia la sabiduría



DEV.F:

Diagrama de Venn

SKILLSETS



Software a Utilizar

Software a ocupar (M-1)

Base:



Colab
Google



MySQL

Complementario:



Anaconda



Jupyter

DEV.E:

220602 Segunda clase

Recursos:

https://drive.google.com/drive/folders/1WY_XfE85tGN_IEkxJFv0AR-pzVi5UbQL

Presentación: https://docs.google.com/presentation/d/11fJ2VEb_SXvf_ce8fSuRQnYi5SdoKaHi/edit#slide=id.p1

220607 tercera clase

Listas, Diccionarios, Booleanos, Condicionales y Ciclos

220609 cuarta clase

Continuacion de clase pasada

For, in, range,

nuevalista = [expresión *for* elemento *in* iterable *if* condicion] #comprehension list

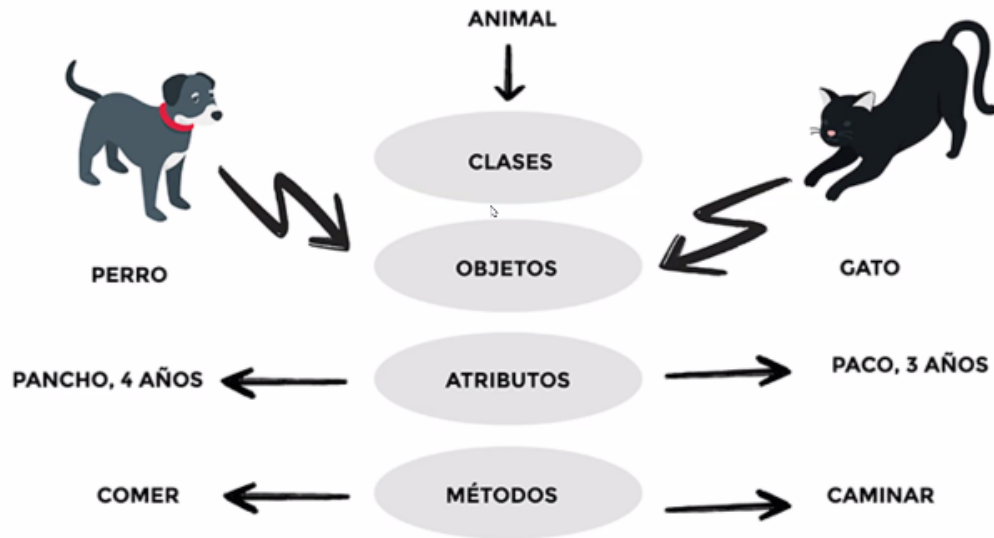
Funciones

def <funcion>(<inputs>):

#codigo

220614 quinta clase

https://docs.google.com/presentation/d/1ZcHx9yZJFW5mNCGWehA8Kj3KeEe8wU9iJQcG39eJSPY/edit#slide=id.gba388653f5_0_6



Objetos

220616 sexta clase

SQL

<https://docs.google.com/presentation/d/1MMA3tbrpCfYT4xFMCIxpGe06NfAW0P81/edit#slide=id.p18>

CREATE

```
CREATE TABLE IF NOT EXISTS
```

```
user
```

```
(email text,  
first_name text,  
last_name text,  
address text,  
age integer,  
PRIMARY KEY (email))
```

INSERT

```
INSERT INTO user
```

```
VALUES
```

```
('bob@example.com',  
'Bob',  
'Codd',  
'123 Fantasy lane  
Fantasy City',  
31)
```

SELECT(READ)

```
SELECT * FROM user ORDER BY age DESC
```

UPDATE

```
ALTER TABLE user ADD COLUMN gender text
```

```
UPDATE user SET gender='M'
```

DELETE

```
DELETE FROM user WHERE email='bob@example.com'
```

DELETE FROM user

DROP TABLE user

AGGREGATE FUNCTIONS

AVG AVG(ALL | DISTINCT)

COUNT

MAX

MIN

SUM

Pendientes CLASES 6 Y 7

REVISAR MEDIA CLASE 6 RESOLVER clase y challenge y revisaxar totalidad de dclase 7 y hacer challenge

220623 Clase 8

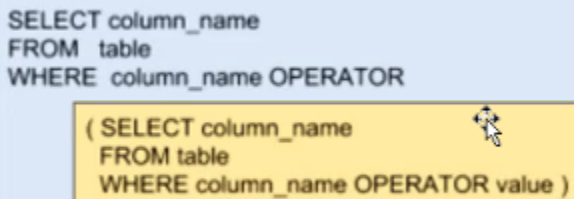
SUBCONSULTAS

Clase: <https://docs.google.com/presentation/d/1E-GELPtfvnFbqmkoTiZdgb1xONxGt1CgHNjXZpiK2HE/edit>

Es una consulta dentro de otra consulta de SQL

Aka subquery o consulta anidada

Puede anidarse dentro de una instrucción SELECT, INSERT, UPDATE o DELETE



The diagram shows a SQL query structure with a subquery. The outer query is highlighted in light blue and consists of the following text:

```
SELECT column_name  
FROM table  
WHERE column_name OPERATOR
```

The inner query, representing the subquery, is highlighted in yellow and is enclosed in parentheses. It contains the following text:

```
( SELECT column_name  
  FROM table  
  WHERE column_name OPERATOR value )
```

A mouse cursor is shown pointing at the closing parenthesis of the subquery.

Modulo 2

220705 Clase 9 INTRO A DATA SCIENCE

¿Como hacer una buena pregunta?

PPT:
https://docs.google.com/presentation/d/1f4nkDRd-UxEiYNU3b7AmPk09aQKdnwSc41aowxA0yos/edit#slide=id.gda890590e2_0_0

Características

- Relevante
- No haya sido respondida anteriormente
- Razonable / probable
- Puede ser respondida con los datos que se cuenta
- Especifica

Tipos de datos

Numeric (Quantitative)

- Continuous (Measured)
 - Airplane speed
 - Time spent waiting in line
- Discrete (Counted)
 - Number of pets
 - Number of packages shipped

Categorical (Qualitative)

- Nominal (Unordered)
 - Married/unmarried
 - Country of residence
- Ordinal (Ordered)
 - ☐ Strongly disagree
 - ☐ Somewhat disagree
 - ☐ Neither agree nor disagree
 - ☒ Somewhat agree
 - ☐ Strongly agree

DEVE:

Genero – Cualitativo Nominal , categorico binario

Introvertido/extrovertido – Cualitativo Nominal, cat binario

Horas de sueño – Cuantitativo Discreto, num cont

Países visitados – Cuantitativo Discreto, num disc

Inseguro 1 al 5 – Cualitativo Ordinal, cat ord

CP – Cualitativo Nominal, cat

Tipos de datos

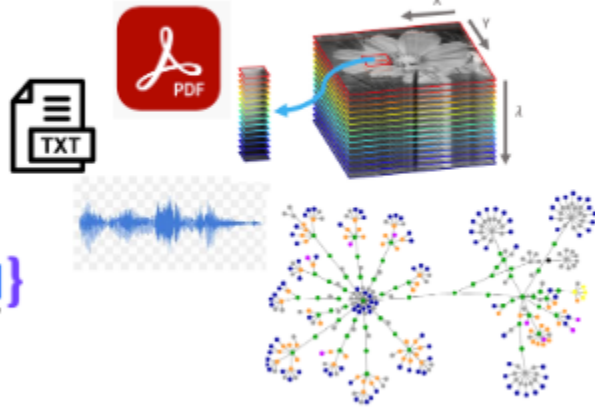
Estructurados:

Stu.	gender	intro_extra	...	dread
1	male	extravert	...	3
2	female	extravert	...	2
3	female	introvert	...	4
4	female	extravert	...	2
...
86	male	extravert	...	3



Semi-estructurados

No estructurados:

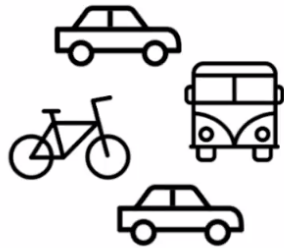


DEV.F.

Estadística descriptiva VS inferencial

Descriptive statistics

- Describe and summarize data



- 50% of friends drive to work
- 25% take the bus
- 25% bike

<https://www.youtube.com/watch?v=4B2xOvKFFz4>

Inferential statistics

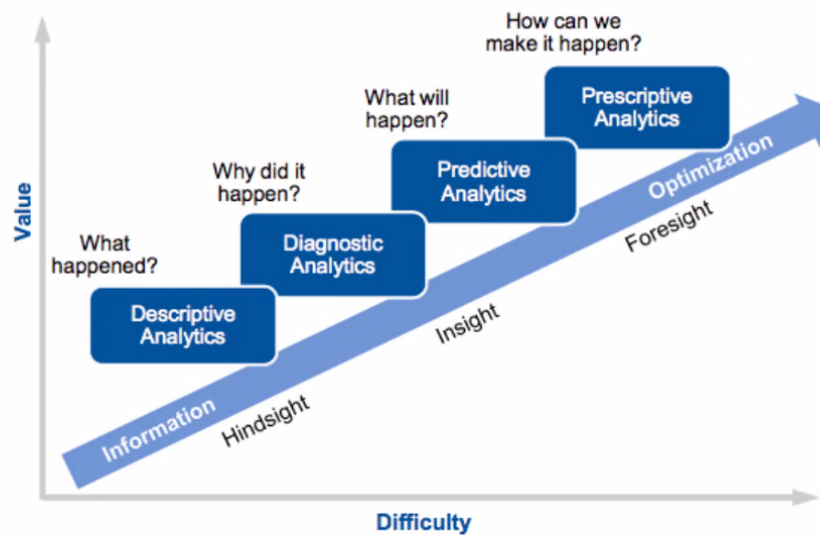
- Use a sample of data to make inferences about a larger population



What percent of people drive to work?

DEV.F.

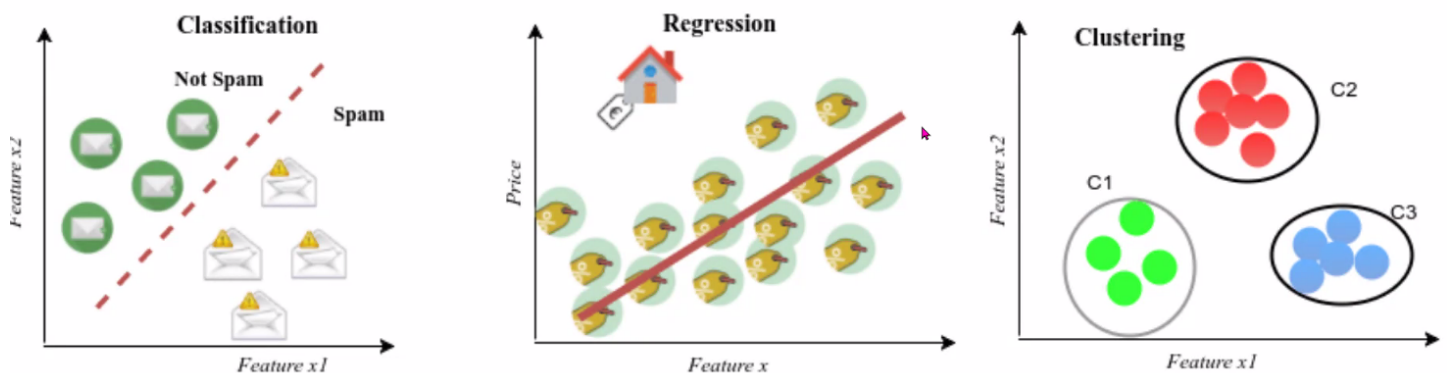
Niveles de Analítica



Source: Gartner (March 2012)

DEV.F

Tareas comunes en data science



DEV.F

Proceso: filtrar, agrupar, colapsar

1. **Filtra:** Plántate un caso particular (datos de un año específico)
2. **Agrupar:** Piensa en una o varias categorías que te permitan agrupar tus datos (género, edad, localidad, etc)
3. **Colapsa:** Piensa en un factor medible que afecte a todos los grupos, una columna numérica de preferencia (total, sumatoria, porcentaje, promedio, etc)

DEV.F

FILTRAR – AGRUPAR – COLAPSAR

- Dos sets de datos, estos datos son reales y fueron captados con algunos sensores de tecnología de tracking de la posición de celulares mediante señales de wifi.

Actividad

Vamos a separarnos en equipos y a generar 3 preguntas por equipo acerca del reto de Klustera. Cada pregunta irá acompañada de la descripción de los 3 pasos para responderla (filtrar, agrupar, colapsar)

Ejemplo: ¿Hay más personas etiquetadas como visitante en fin de semana?

1. **Filtrar:** Obtener los registros en días sábado y domingo (day_tz)
2. **Agrupar:** Visitantes y no visitantes (visitor)
3. **Colapsar:** Contar cuántos registros hay en cada grupo (count)

```
datae= 'ruta.csv'
```

```
dte = pd.read_csv(datae,index_col = 0)
```

```
import pandas as pd
```

From Natanael German Ramirez to Everyone 06:35 PM

gracias

From Alfonso Miguel to Everyone 06:35 PM

read_csv

PENDIENTE: TRABAJAR CON FINALIZACION DE CLASE ACTIVIDADES Y EJEMPLO

220707 Clase 10

Colección y Adquisición de Datos en Python

PPT:

https://docs.google.com/presentation/d/1mKYE_B8596NQeXkDtxVxQjYPmQPJIK_ikUPInMzWJu4/edit#slide=id.gba388601b5_0_12

Drive: <https://drive.google.com/drive/folders/1ovHCmBt2EDbl8PjLriZQftjAW3DVaxX0>

220712 Clase 11

Data Manipulation

NumPy

PPT:

https://docs.google.com/presentation/d/1YkS-yBV6J9gqj4QCTZfPw0nbLnXcBcw7ojGTE-GX3Q/edit#slide=id.gba388601b5_0_141

220714 Clase 12

Data Modeling

Tidy Data

PPT:

https://docs.google.com/presentation/d/1pzXSc61nym-3-Dog4VnFiShHHJT8Alrhhs8M0iWuHs/edit#slide=id.gba388653f5_0_6

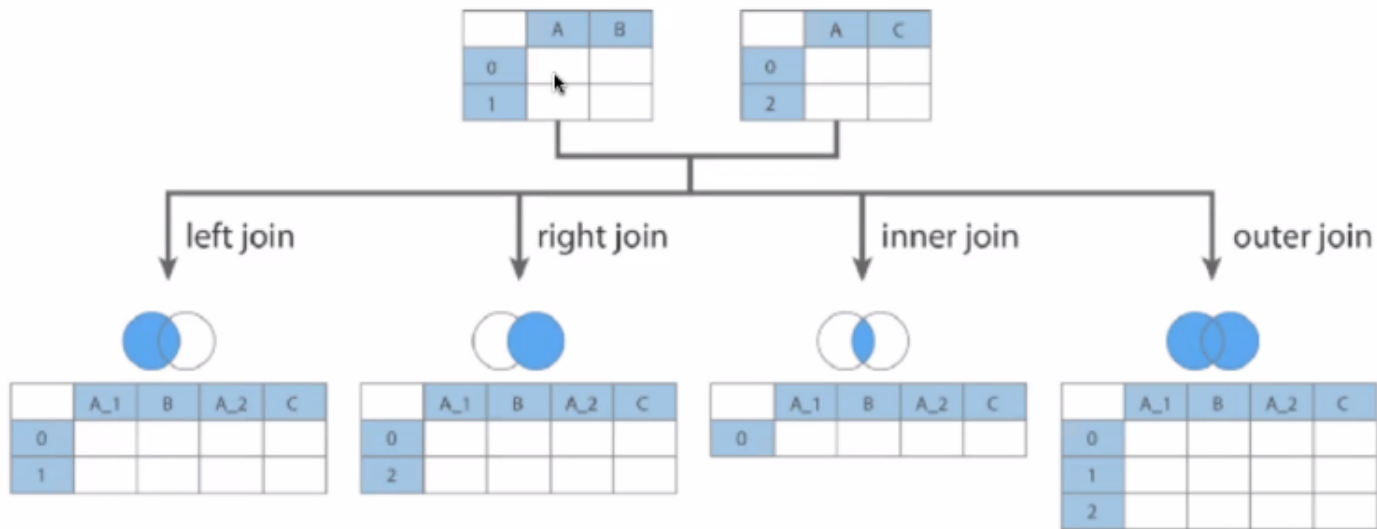
Metodos de Pandas

`Pd.melt(df)` ; Gather Columns in t rows

`Df.pivot(columns='var', values='val')` ; Spred Rows into columns

`Pd.concat([df1,df2])` ; Append rows of DataFrames

`Pd.concat([df1,df2],axis=1)` ; Append columns of Dataframes



Colab: <https://colab.research.google.com/drive/1uC0DsKltjv9k0fPb3nFgwM-avQXUOcgE?usp=sharing>

Data:

Religion, <https://drive.google.com/file/d/1LICSGqQmAQMn6OilfD3AupT9ool3O0gi/view>

Wider, <https://drive.google.com/file/d/1xZA9kliuum0JILAvQz3o489vUMG2mTVx/view>

Unite, <https://drive.google.com/file/d/16BpBkUwjuaW1ZISJ366GJ0AUQdlxELhZ/view>

Separate, <https://drive.google.com/file/d/16BpBkUwjuaW1ZISJ366GJ0AUQdlxELhZ/view>

Challenge 1: <https://colab.research.google.com/drive/1tCk5fJcj4Y3RsVaCr5leYIIH3hZPE6I2#scrollTo=jV45G9Dd2dFZ>

Challenge 2: <https://colab.research.google.com/drive/1TMX4bIlnUwo67N-KjEDcENADBXPvHb7L>

Data: https://drive.google.com/drive/folders/1gVPi_M8h98oJciODyKRnKBQv-FSFjhY

Resumen

Pandas Tidy Data

`pandas.melt()`

```
pd.melt(df, id_vars='pais', var_name='año', value_name='Cantidad')
df2.melt(id_vars='religion', var_name='ingreso', value_name='frecuencia')
```

`dataframe.pivot_table()`

```
df_wider.pivot_table(values='cuenta', index=['pais', 'anio'], columns='tipo')
```

`dataframe.col.str.split()`

```
df_separate.tasa.str.split("/", expand=True)
```

```
dataframe.drop()
df_separate.drop(columns='tasa')

dataframe.concat([df,df2],axis=1)
pd.concat([df_separate_drop,df_separate_col],axis=1)
pd.concat([df_unite.drop(columns=['siglo','anio']),df_unite['siglo']*100 + df_unite['anio']],axis=1)
```

Merge Methods

Merge, inner, left right outer

`pandas.merge(left_df , right_df ,how: str='Inner',...)`

how:{'left','right','outer','inner'} on: label or list

El argumento how del método merge especifica cómo determinar qué variables (columnas) serán incluidas en la tabla resultante. Si una valores no aparece en las tablas de la izquierda o de la derecha, los valores de la tabla combinada serán NA. Aquí hay un resumen de las opciones de how y sus nombres equivalentes de SQL:

Merge method	Nombre del SQL Join	Descripción
left	LEFT OUTER JOIN	Usa únicamente las variables de la izquierda
right	RIGHT OUTER JOIN	Usa únicamente las variables de la derecha
outer	FULL OUTER JOIN	Utiliza la unión de las variable de ambos data frames
inner	INNER JOIN	Utiliza la intersección de las variables de ambos data frames

Important Methods **Pandas** Packages

@TheRavItShow

Data Importing

```
pd.read_csv()  
pd.read_table()  
pd.read_excel()  
pd.read_sql()  
pd.read_json()  
pd.read_html()  
pd.DataFrame()  
pd.concat()  
pd.series()  
pd.date_range()
```

Data Cleaning

```
pd.fillna()  
pd.dropna()  
pd.sort_values()  
pd.apply()  
pd.groupby()  
pd.append()  
pd.join()  
pd.rename()  
pd.to_csv()  
pd.set_index()
```

Data Statistic

```
pd.head()  
pd.tail()  
pd.describe()  
pd.info()  
pd.mean()  
pd.median()  
pd.count()  
pd.std()  
pd.max()  
pd.min()
```

220719 Clase 13

Data Cleaning - Limpieza de Datos con Python

Dates

Librerias

```
import pandas as pd
import numpy as np
import seaborn as sns
import datetime
import chardet
```

Ejemplo de cambio de fecha

Cuando tenemos fechas así, 0 3/2/07 1 3/22/07 2 4/6/07 3 4/14/07 4 4/15/07

Las convertimos de esta manera

```
df['date_parsed'] = pd.to_datetime(df['date'], format='%m/%d/%y')
df['date_parsed'].head()
```

Y para mostrar por partes es con estos métodos

```
.dt.day,
.dt.month,
.dt.year
```

De esta manera limpiamos con suma de null en true y drop de nans

```
df.isnull().sum()
df = df.dropna()
```

Se grafica el histograma de l a siguiente manera

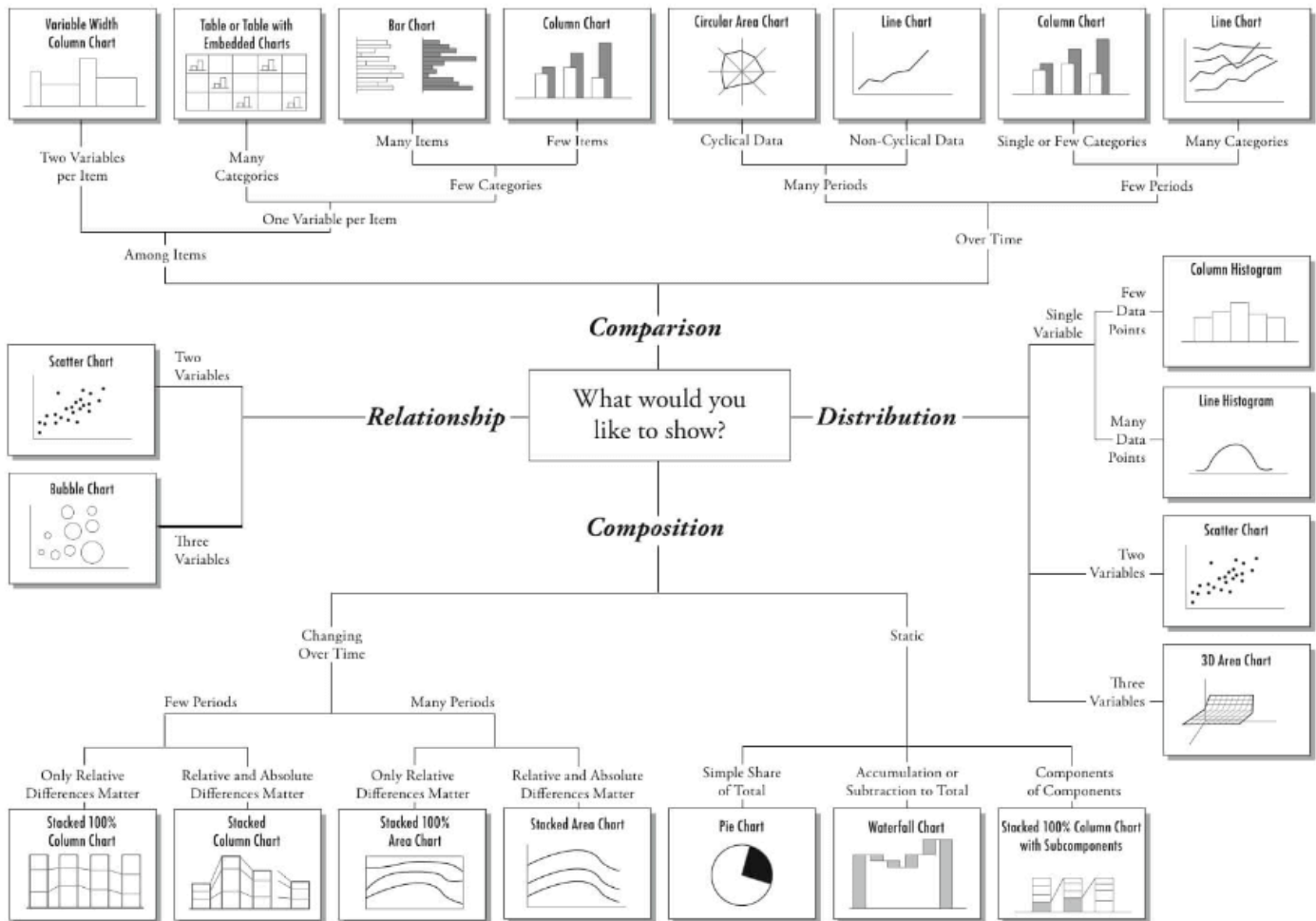
```
sns.histplot(df, bins=31)
```

adasd

220726 Clase 14

Data Visualization - Visualizacion de datos con Matplotlib + Seaborn

Chart Suggestions—A Thought-Starter



Proyecto: <https://cogitovsmachina.github.io/codelabs/el-deportista-mejor-pagado/#0>

[220801 Comlemento Clase 14](#)

[220802 Clase CRISP](#)

Presentacion

Presentaciones

Cada equipo tendrá 10-15 minutos para presentar su proyecto

De preferencia, solo un integrante del equipo presentará pero los demás integrantes pueden complementarle.

Al finalizar tu presentación, por favor, responde lo siguiente:

- ¿Cuáles fueron sus mayores logros al realizar este reto?
- ¿Cuáles fueron los mayores retos técnicos a los que se enfrentaron?
- ¿Cuál fue el mayor reto al que se enfrentaron trabajando en equipo?



DEV.F.

Modulo 3 Intro a Fundamentos

220809 Clase 15

Sesion 1 .- Intro a Calculo

Evaluacion Diagnostico: <https://forms.gle/GvEYDAKXww8CzNWj7>

Presentacion_

<https://docs.google.com/presentation/d/1-GoUfUd65iVagrlzPlbV8s7vhjIMd3i0/edit?usp=sharing&ouid=104062509548216645094&rtpof=true&sd=true>

- Prestar atención en la composición, slide 23 ejemplo de desarrollo de $g(f(x))$
- Formulario: <https://drive.google.com/file/d/1EBoka0iTktlvUJ1AIBQ7eyx2mwo0b2dP/view>
- Prestar atención en regla de la cadena, slide 33, derivada de una composición
- Criterio para evaluar máximos y mínimos:
https://drive.google.com/file/d/1zpW_gsOKzeolYonrOtPG-z9mIF-JbbZQ/view

220811 Clase 16

Sesion 2.- Calculo para Machine Learning

Presentacion_

<https://docs.google.com/presentation/d/18twdBxyUFUdKan15gaKqPa88iNv2DN8m/edit?usp=sharing&ouid=107142877221780888677&rtpof=true&sd=true>

- Prestar Atencion en Funcion de perdida(Loss Function) slide 20

220816 Clase 17

Sesion 3.- Algebra Lineal para Machine Learning

Se trabaja con el notebook:

<https://colab.research.google.com/drive/16ZPKoFK0CC54pRRK6btdb03xqFypITiM?usp=sharing>

Vectores y Escalares

- Escalar = λ es un numero
- Multiplicacion de vector y escalar \Rightarrow vector
- Producto punto \Rightarrow multiplicación de dos vectores que resulta en un escalar(numero), es lo mismo $a \cdot b$ que $b \cdot a$
- Producto cruz \Rightarrow operación entre dos vectores que da un vector perpendicular al plano de ambos vectores, $a \times b$ no es necesariamente lo mismo que $b \times a$
- Fundamentos de la algebra lineal
 - Estos vectores unitarios son independientes entre sí, es decir, que no es posible formar uno a partir de otro(s).
 - Generan todo el espacio (para 3d) o todo el plano xy (para 2d) al combinar linealmente (es decir estirándolos y sumándolos) dichos vectores.
- Las bases i, j, k son llamadas bases canónicas.

Matrices

Las transformaciones lineales cumplen con las siguientes propiedades:

Transforman rectas a rectas, no se pueden curvar.

Mantiene el origen en la misma posición.

Mantiene la misma distancia entre los puntos de la cuadrícula.

Transformación: Forma Sofisticada de decir Función

De manera numérica, la forma de representar una transformación es mediante una matriz, donde las columnas de una matriz representan el punto donde termina cada vector unitario de la base al aplicar la transformación.

Suma de Matrices = Deben de ser de las mismas dimensiones

Multipliación de matrices:

Matriz por escalar => resulta en matriz

Matriz por vector => resulta del producto punto

Matriz por matriz => resulta en matriz de dimensiones filas A por columnas B

OJO LA FUNCION MATRIX, NO NOS VA A PERMITIR USAR OTRO TIPO DE DATO QUE NO SEA NUMERICO, POR ELLO DEBE DE CUIDARSE CUANDO USAR NP.MATRIX O NP.ARRAY

Matriz Traspuesta => se transponen los elementos filas por columnas

Determinante => escalar que se relaciona con el área o volumen del plano o subespacio generado por las columnas de la matriz

Si el det es diferente de 0 indica que los vectores columna son independientes

Si el det es 0, indica que alguno de los vectores columna es dependiente.

Matriz Inversa

Una matriz puede o no tener inversa; si **no tiene inversa decimos que es una matriz singular** 🤖

Una matriz singular implica que no tiene inversa, su determinante es 0 y sus vectores columna son linealmente dependientes

No singular es que **sí tiene inversa**, su **determinante es diferente de 0** y todos sus **vectores columna son linealmente independiente**

Al **número de vectores linealmente independientes** en una matriz, también se le llama **rango o dimensión**.

La matriz inversa se utiliza para calcular el escalar de una ecuación, por ejemplo:

$$A \cdot x = b$$

$$x = A^{-1} \cdot b$$

Por ello es importante saber si una matriz es singular para así saber si hay respuesta a las ecuaciones lineales

Matriz Cuadrada

Matriz Rectangular

Matriz Simétrica

Matriz Triangular,

Triangular superior, np.triu(Matrix)

Triangular inferior, np.tril(Matrix)

Matriz Diagonal, hay que tener cuidado con esta ya que al calcularla en python así, $z = \text{np.diag}(\text{Matrix})$ al asignarlo a z va a regresar solo un array lineal, dado que el triangular superior e inferior es igual a cero

220818 Clase 18

Sesion 4.- Algebra Lineal para Machine Learning, segunda parte

Presentación: <https://docs.google.com/presentation/d/1XYFFxvsgsmm0lYKxU3ffBoCD2D40gpBB/edit#slide=id.p1>

Solución de mult de matriz: <https://drive.google.com/file/d/1l6EE8l0fuOlGzFQ2DFHlVMxF1Y78OeOW/view>

Tipos de Matrices

Matrices Cuadradas

Matrices Rectangulares

Matriz Simetrica

Matriz Triangular,

Triangular superior, np.triu(Matrix)

Triangular inferior, np.tril(Matrix)

Matriz Diagonal, hay que tener cuidado con esta ya que al calcularla en python así, $z = \text{np.diag}(\text{Matrix})$ al asignarlo a z va a regresar solo un array lineal, dado que el triangular superior e inferior es igual a cero

Ortogonal, $A^T = A^{-1}$ inversa e igual a transpuesta, por lo que $A \cdot A^t = A^t \cdot A = I$

Ortonormal,

Matrices como representación de ecuaciones

$$\begin{aligned}x + 2y + z &= 8 \\2x + 2y + z &= 9 \\3x + 3y + 5z &= 24\end{aligned}$$

$$\begin{pmatrix} 1 & 2 & 1 \\ 2 & 2 & 1 \\ 3 & 3 & 5 \end{pmatrix} * \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 8 \\ 9 \\ 24 \end{pmatrix}$$

$$A * X = B$$

Factorizaciones

Factorización $A = C.R \Rightarrow$ Matrix.rref()

Row reduce echelon form, nos ayuda a visualizar cantidad de columnas independientes

PRESENTACION Gilbert Strang MIT:

<https://www.youtube.com/watch?v=aazrfdysfi0&list=PLUI4u3cNGP61iQEFiWLE21EJCxwmWvvek>

$$A\mathbf{x} = \begin{bmatrix} 1 & 4 & 5 \\ 3 & 2 & 5 \\ 2 & 1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} x_1 + \begin{bmatrix} 4 \\ 2 \\ 1 \end{bmatrix} x_2 + \begin{bmatrix} 5 \\ 5 \\ 3 \end{bmatrix} x_3$$

= linear combination of columns of A

Basis for the column space / Basis for the row space

Include column 1 = $\begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$ in C Include column 2 = $\begin{bmatrix} 4 \\ 2 \\ 1 \end{bmatrix}$ in C

DO NOT INCLUDE COLUMN 3 = $\begin{bmatrix} 5 \\ 5 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} + \begin{bmatrix} 4 \\ 2 \\ 1 \end{bmatrix}$
IT IS NOT INDEPENDENT

$$A = CR = \begin{bmatrix} 1 & 4 \\ 3 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \quad \begin{array}{l} \text{Row rank} = \\ \text{column rank} = \\ r = 2 \end{array}$$

The rows of R are a basis for the row space

Basis for the column space / Basis for the row space

Include column 1 = $\begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$ Include column 2 = $\begin{bmatrix} 4 \\ 2 \\ 1 \end{bmatrix}$

DO NOT INCLUDE COLUMN 3 = $\begin{bmatrix} 5 \\ 5 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} + \begin{bmatrix} 4 \\ 2 \\ 1 \end{bmatrix}$
IT IS NOT INDEPENDENT

Basis has 2 vectors A has rank $r = 2$ $n - r = 3 - 2 = 1$

Counting Theorem $A\mathbf{x} = \mathbf{0}$ has one solution $\mathbf{x} = (1, 1, -1)$

There are $n - r$ independent solutions to $A\mathbf{x} = \mathbf{0}$

$A = CR$ shows that column rank of A = row rank of A

1. The r columns of C are independent (by their construction)
2. Every column of A is a combination of those r columns (because $A=CR$)
3. The r rows of R are independent (they contain the r by r matrix I)
4. Every row of A is a combination of those r rows (because $A = CR$)

Key facts

The r columns of C are a **basis** for the column space of A : **dimension r**

The r rows of R are a **basis** for the row space of A : **dimension r**

Factorización $A=LU$

$$A\mathbf{x} = L\mathbf{U}\mathbf{x} = \mathbf{b}, \quad A\mathbf{y} = L\mathbf{U}\mathbf{y} = \mathbf{b}$$

1. Primero, resolvemos $L\mathbf{y} = \mathbf{b}$ para \mathbf{y}
2. Segundo, resolvemos $\mathbf{U}\mathbf{x} = \mathbf{y}$ para \mathbf{x}

Siendo, L la matriz triangular inferior

U la matriz triangular superior

$P, L, U = \text{scipy.linalg.lu}(A) \Rightarrow$ aquí se obtienen tres valores P (Matriz Identidad), Lower y Upper

Para Inversas:

$$A^{-1} = U^{-1}L^{-1}P. \Rightarrow \text{np.linalg.inv}(A)$$

Factorización A=QR # AQUÍ VOY EN EL REPASO

Presentación siguiente: https://drive.google.com/file/d/1MoX_WV3-c0bTy5Q3OynjzbZEgoRsIPbp/view?usp=sharing

Valores y Vectores Propios

Vector propio.- Es aquel que cuando se la aplica un producto punto entre la matriz y el vector v solo se estira o se permanece igual(no se mueve de s dirección, este vector v es propio de la matriz A, nota: SOLO APLICA EN MATRIZ CUADRADA

Se determina si se cumple lo siguiente, $\det(A - \lambda I) = 0$

Eigen Value Desscomposition es para matrices cuadradas

`linalg.eig`

Single Value Descomposition, es para matrices no cuadradas

$A = U(\text{Sigma})V$

TERMINAR CLASE

220823 Clase 19

Sesion 5.- Estadística y Probabilidad

Presentación:

https://docs.google.com/presentation/d/1duOjBkBTjwsNMdEOgwfzozEDHsXeFAtw/edit#slide=id.g11532e52183_0_0

Diagramas de Venn, **hacerlos acorde ejercicio slide 9**

Evento o suceso, es un subconjunto de un espacio muestral.

Permutacion, nos interesa el orden(no repetitivo) r objetos seleccionados de n objetos distintos

$$nPr = n(n - 1)(n - 2) \dots (n - r + 1)$$

Combinacion, no nos interesa el orden, pueden combinarse(duplicidad)

$$nCr = nPr/r!$$

$$\binom{n}{r} = \frac{n(n-1)(n-2)\dots(n-r+1)}{r!}$$

Probabilidad Enfoque Clasico

Si existen **n posibilidades igualmente probables**, una de las cuales debe ocurrir y **s considerarse como favorables** o como éxito, entonces la probabilidad de un “éxito” está dada por s/n .

Si 3 de 20 neumáticos almacenados son defectuosos y se seleccionan aleatoriamente para su inspección 4 de ellos (equiprobables), ¿cuál es la probabilidad de que sólo sea incluido uno de los neumáticos defectuosos?

3C1 => Elegir 1 defectuoso de los 3 defectuosos

17C3 => Elegir 3 buenos de los 17 buenos

20C4 => elegir 4 de los 20 existentes

$E/S = ((3C1)*(17C3))/(20C4) = 0.42$

Probabilidad Enfoque Clasico

La probabilidad de un evento o resultado es la proporción de veces en las que el evento ocurrirá en una repetición prolongada de experimentos.

Axiomas de Probabilidad

Expliquemos lo que entendemos por probabilidad de un evento.

Dados un espacio muestral finito S y un evento A en S, definimos $P(A)$, la probabilidad de A satisface las siguientes tres condiciones:

Axioma 1. $0 \leq P(A) \leq 1$ para cada evento A en S

Axioma 2. $P(S) = 1$

Axioma 3. Si A y B son cualesquiera eventos mutuamente excluyentes en S, entonces

$$P(A \cup B) = P(A) + P(B)$$

Teoremas Elementales

Regla de cálculo de probabilidad de un evento

Regla General de la Adición

Probabilidad de complemento

Probabilidad condicional e Independencia

Definición: Sean A y B dos eventos asociados a un experimento. Denotaremos $P(B | A)$ la probabilidad condicional de B dado que A ha ocurrido.

CALCULO EJERCICIO Slide 39

Revisar sucesos independientes y dependientes,

EJERCICIOS SLIDE 42

220825 Clase 20

Sesion 6.- Estadística y Probabilidad parte 2

Presentación:

https://docs.google.com/presentation/d/18_cFtqXCGM4fNHqgTTe6Y0nXQqsnzjFu/edit?usp=sharing&oid=107142877221780888677&rtpof=true&sd=true

Empezamos de como resolver la paradoja de cumpleaños

Repaso clase pasada

PROBABILIDAD CONDICIONAL

DOS EJERCICIOS

TEOREMA DE LA MULTIPLICACION PARA PROBABILIDAD CONDICIONAL(DOS EVENTOS INDEPENDIENTES, ESTO POR QUE NO HAY INTERSECCION ENTRE AMBAS)

PROCESO ESTOCASTICO(FINITO), DIAGRAMA DEL ARBOL

Teorema de Bayes

Distribución aleatoria

REREREPASAAAAARRRRR QUE NO PUSE ATENCION

220830 Clase 21

Sesion 7.- Estadística parte 1

Presentación:

<https://docs.google.com/presentation/d/13EYylwQ5ITWmuxTT7w4eF2hH-IMyH6-/edit?usp=sharing&oid=107142877221780888677&rtpof=true&sd=true>

Estadística Descriptiva, recoge , almacena, ordena, grafica,

Estadística Inferencial, concluye apartir de una muestra

Aprendizaje Estadisitco(statistical learning)

Poblacion y Muestra

Bias, aleatoriedad

Métricas de Tendencia Central

Media, valor promedio

Mediana, numero medio en el conjunto, promedio de los dos números medios

Moda, valor con mayor frecuencia

Tipos de Distribución

Sesgada por la izquierda

Simetrica

Sesgada por la derecha

.
.
.

Correlations

Notebook: https://drive.google.com/file/d/1FZJ_py01AxR6bsuVy_dl4vb2Ljh1itUs/view

Datasets: <https://drive.google.com/drive/folders/173bnPDixrf3hymTYn6lFsff92cl137xK?usp=sharing>

Challenge:

Tarea (Opcional)

- Realizar todo el Notebook, resolver las tareas del notebook y poner comentarios personales sobre lo que se pregunta.
- Deadline: 07 de agosto de 2022
- ¿Qué ganan? Conocimiento...y el tener retroalimentación directa y personalizada sobre su trabajo...¿no es suficiente motivación? además de una recomendación en LinkedIn por parte de los senseis del curso.

DEV.F.

Sesión 8.- Estadística parte 2

Presentación:

https://docs.google.com/presentation/d/1ikA8fL1J_pEpMxayiyBWfVig1WImO2-2/edit?usp=sharing&ouid=107142877221780888677&rtpof=true&sd=true

Inferencia

Normalización de Distribuciones

Intervalo de Confianza

Menor a 30, T de Student

Mayor igual a 30, Normal Estándar

Pruebas de Hipotesis

Hipotesis nula H_0 e Hipotesis Alternativa H_1 , ambas son enunciados mutuamente excluyentes

Significancia Estadística y P Values

Nivel de significación(alfa), limite para juzgar resultado como estadísticamente significativo

Hipotesis Bilateral o unilateral

Estadístico que se respeta sabe que:

- El p-value obtenido refleja la probabilidad de rechazar la H_0 siendo esta verdadera; **en ningún caso prueba que la hipótesis alternativa, de que si hay efecto o diferencia, sea verdadera.**
- Es decir, rechazar la hipótesis nula no implica que la alternativa sea verdadera.
- Por ello siempre decimos "No hay evidencia suficiente para rechazar H_0 " ó "Hay evidencia suficiente para rechazar H_0 ".

A/B Testing

Data: <https://drive.google.com/drive/folders/10ccmHp6aH190bLLCdf9sCdjKZ-FdPGj4>

Colab: <https://colab.research.google.com/drive/14lC8v6qSujgWCiF50CtOAWaurOdXpznz>

MODULO 4 DATA WRANGLING

220912 Clase 23

Senseis

Mali, Data Engineer

Miguel, Super computadoras, UNAM

Sesión 1.- Introducción a la Ingeniería de Datos

Ppt:

https://docs.google.com/presentation/d/1qQ-32Grg_EzXE8oEs_kUDME52qg3DWH7xvOfecoKdE/edit#slide=id.gfc0815af54_1_116

Drive Sensei: <http://bit.ly/mdsai-gdrive>

Lecturas:

<https://medium.com/@rchang/a-beginners-guide-to-data-engineering-part-i-4227c5c457d7>

<https://www.altexsoft.com/blog/datascience/what-is-data-engineering-explaining-data-pipeline-data-warehouse-and-data-engineer-role/>

Notebook: <https://colab.research.google.com/drive/1QvVOJKYG6wi5gdxpkAlezquXCqljP7t>

Challenge Examen de Practiva GCP:

https://docs.google.com/forms/d/e/1FAIpQLSfkWEzBCP0wQ09ZuFm7G2_4qtkYbfmk_0getojdnPdCYmq37Q/viewform

220913 Clase 24 clase grabada por 15 de sep

Clase video:

https://devf-mx.zoom.us/rec/share/c0wl-0_5lqZFW-r5SzRiWMATQGB7HITNGBSrS_CP_COAIUBUIBHpNENPZpIQpoD.Q3r-bBD6feno9dE1?startTime=1663206168000

Código de acceso: 8bam8?Cj

PPT, Notebook y Challenge, estan en Campus

SQLite Síntaxis

<https://sqlite.org/lang.html>

<https://www.sqlite.org/syntax.html>

CRUD SQL

220920 Clase 25

220922 Clase 26

Sesion 4.- APIs

PPT: <https://docs.google.com/presentation/d/1y3Din8D19kqNVpar4aAZRuMfggVKC2JG-DvE0gMAPGA/edit>

Notebook: <https://colab.research.google.com/drive/1y8okA5Jz0plcbX6a99lSkRli1Gcz5npd?usp=sharing>

Challenge: https://colab.research.google.com/drive/1llgc3H-MjRe0TtPY5anIRL4rrY_7eCax

220927 Clase 27

Sesion 5.- Data Wrangling y Automatización con CLI

PPT:

Data Wrangling;

https://docs.google.com/presentation/d/1lGsT2aoKTPVsM1uRwpEOg78h3eNro31BFnHMaX5KdZw/edit#slide=id.gd7e9e8ef98_0_281

Postman:

https://docs.google.com/presentation/d/1x3W1XKhC-K9ODxIUJZroPc2UOtp_Q0-DnY8CaMO5z30/edit#slide=id.g118f5939043_1_282

Command Line:

<https://colab.research.google.com/drive/1MEm2OtcbsnhxMFqXP1d4-6UKRSXi-wNZ#scrollTo=YZ7nKJaVY7XI>

Challenge: https://colab.research.google.com/drive/1bEOHzSCWU8W_upniP4pkGKnfYWts_asj

221004 Clase 28

Sesion 6.- Web Scrapping

PPT

Clase Mali:

<https://docs.google.com/presentation/d/1tna71VqQ9el1w6B79JJN-Lgb8LdbuKK7wNDFfr2VqQ0/edit#slide=id.p41>

Clase Campus:

https://docs.google.com/presentation/d/1GJ7j4DyUmHDbEzZlIfGLmC-YROpslNXxs_0B9OKt-6M/edit#slide=id.gafe96b6587_0_212

Datasets y Notebooks Selenium: <https://drive.google.com/drive/u/0/folders/19Y050vLC89kaqHrV4C1gJ5VTBHRwC7cn>

Como correr Jupyter en VScode: <https://code.visualstudio.com/docs/datascience/jupyter-notebooks>

Challenge:

