

TRABAJO PRÁCTICO 1 – PANDAS Y VISUALIZACION DE DATOS

Nombre y Apellido: Mauricio Rodríguez Bertella

Padrón: 100624

Ayudante Asignado: Nati

Primera Parte - Pandas

Consulta 1: 8 -

https://github.com/MauRodriguez/tp1.datos.2021.2/blob/master/consultas/consulta_8.ipynb

CONSIDERACIONES: Conviene trabajar con timedelta, por ejemplo, con la diferencia a la fecha actual.

Consulta 2: 9 -

https://github.com/MauRodriguez/tp1.datos.2021.2/blob/master/consultas/consulta_9.ipynb

Consulta 3: 17 -

https://github.com/MauRodriguez/tp1.datos.2021.2/blob/master/consultas/consulta_17.ipynb

Consulta 4: 18 -

https://github.com/MauRodriguez/tp1.datos.2021.2/blob/master/consultas/consulta_18.ipynb

CONSIDERACIONES: Pregunta cuál es el bloque con menos, pero hay un montón que no tiene ninguna. Alcanza con mostrar una tabla donde se vean claramente los “buckets” con ceros.

Consulta 5: 33 -

https://github.com/MauRodriguez/tp1.datos.2021.2/blob/master/consultas/consulta_33.ipynb

Consulta 6: 40 -

https://github.com/MauRodriguez/tp1.datos.2021.2/blob/master/consultas/consulta_40.ipynb

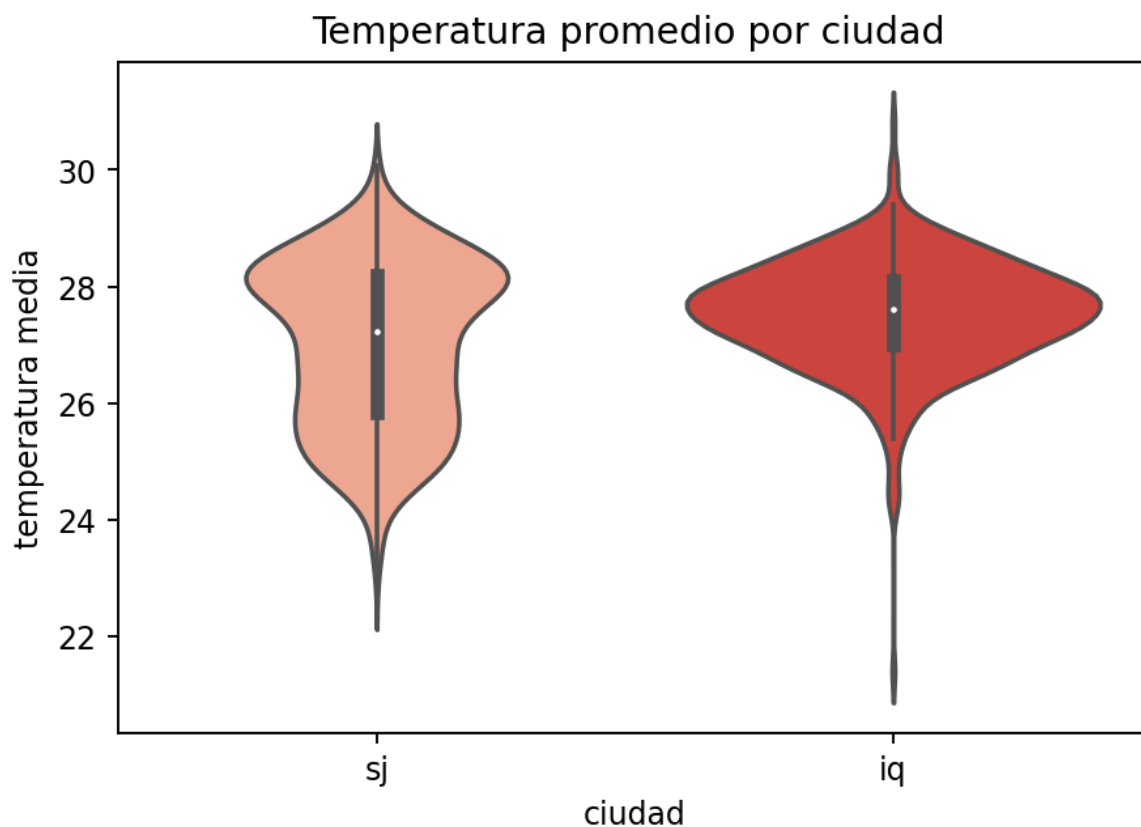
Consulta 7: 43-

https://github.com/MauRodriguez/tp1.datos.2021.2/blob/master/consultas/consulta_43.ipynb

CONSIDERACIONES: Conviene trabajar con timedelta. Para juntar por username/ip se puede hacer `df[user'].fillna(df[ip'])`. Hay que hacer más de un groupby porque si no el merge queda de mucho tamaño y contraproducente trabajar con tanta memoria con tal de hacer 1 solo groupby.

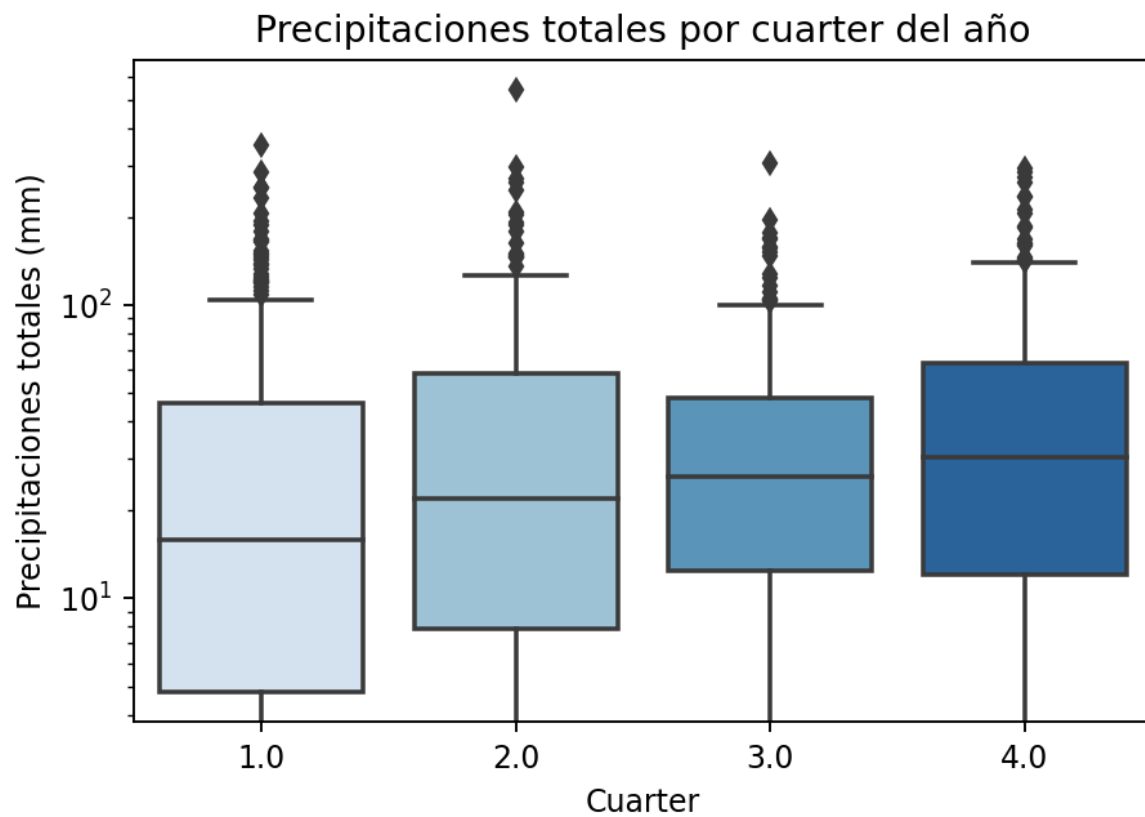
Segunda parte - Visualización de datos

Visualización 1: ViolinPlot – DengAI: Predicting Disease Spread –



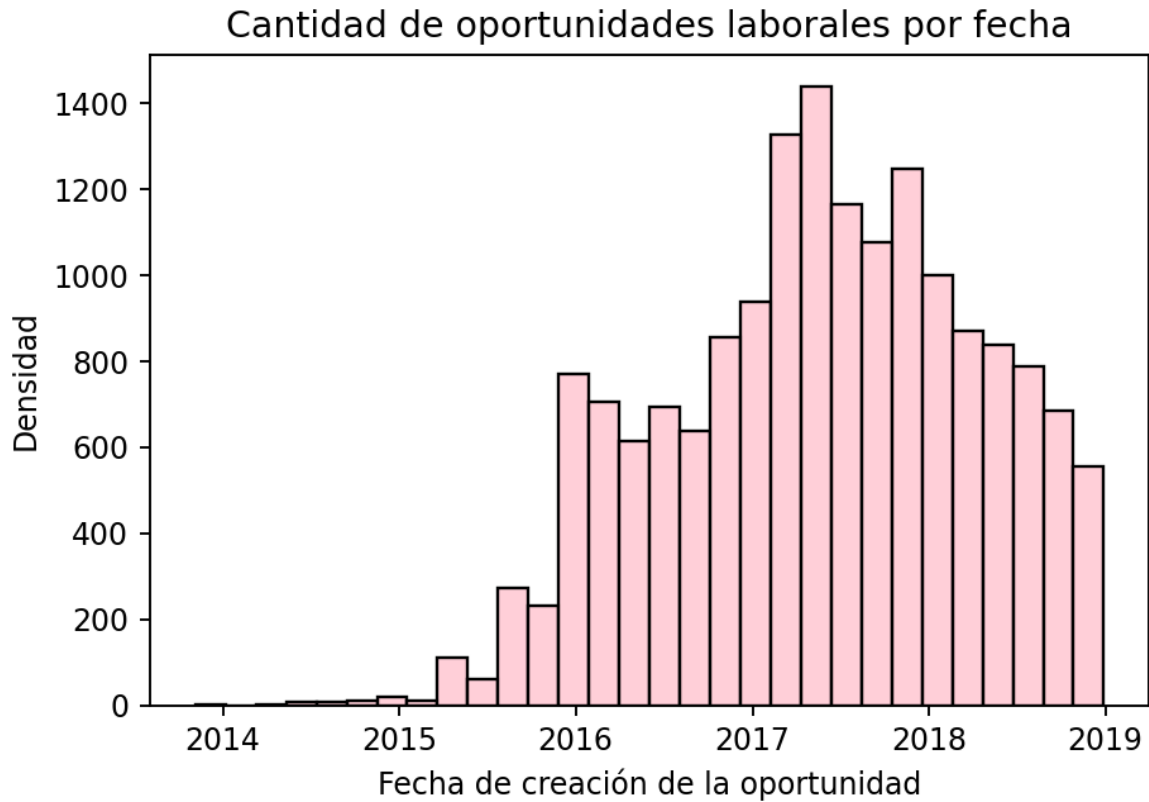
<https://github.com/MauRodriguez/tp1.datos.2021.2/blob/master/visu/visu-dengue.ipynb>

Visualización 2: BoxPlot – DengAI: Predicting Disease Spread –



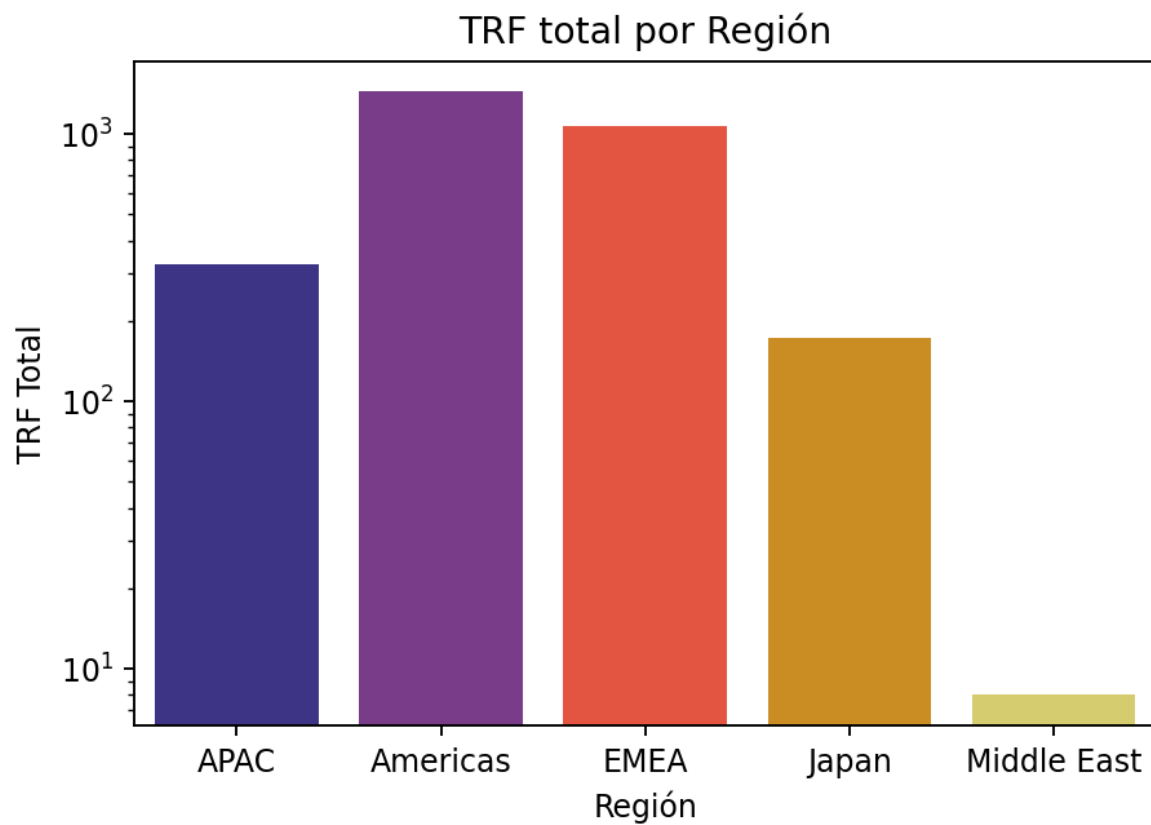
<https://github.com/MauRodriguez/tp1.datos.2021.2/blob/master/visu/visu-dengue-2.ipynb>

Visualización 3: HistogramPlot – Predicción de éxitos en oportunidades comerciales –



<https://github.com/MauRodriguez/tp1.datos.2021.2/blob/master/visu/visu-laboral.ipynb>

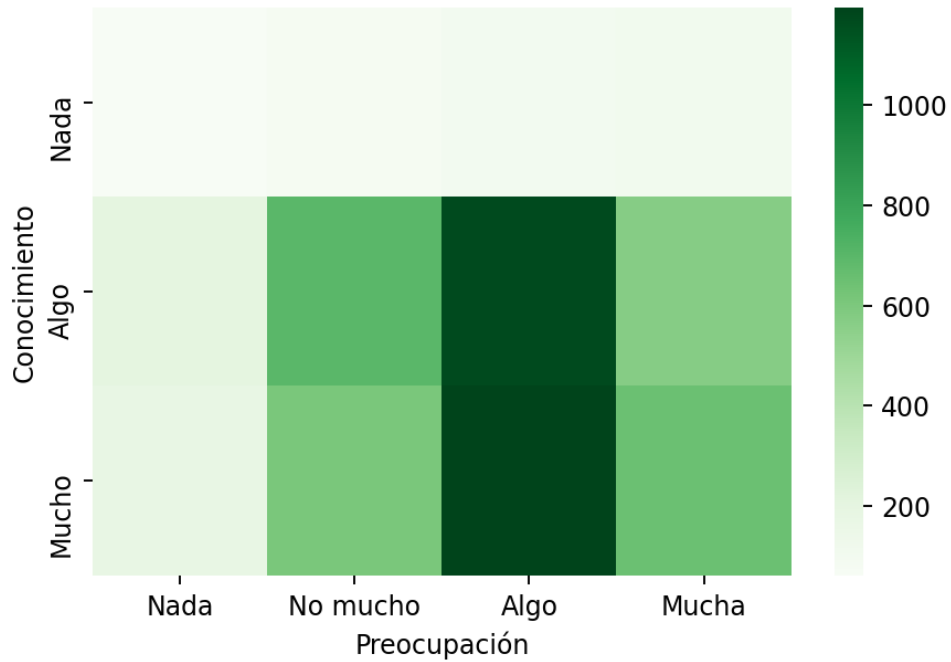
Visualización 4: BarPlot – Predicción de éxitos en oportunidades comerciales –



<https://github.com/MauRodriguez/tp1.datos.2021.2/blob/master/visu/visu-laboral-2.ipynb>

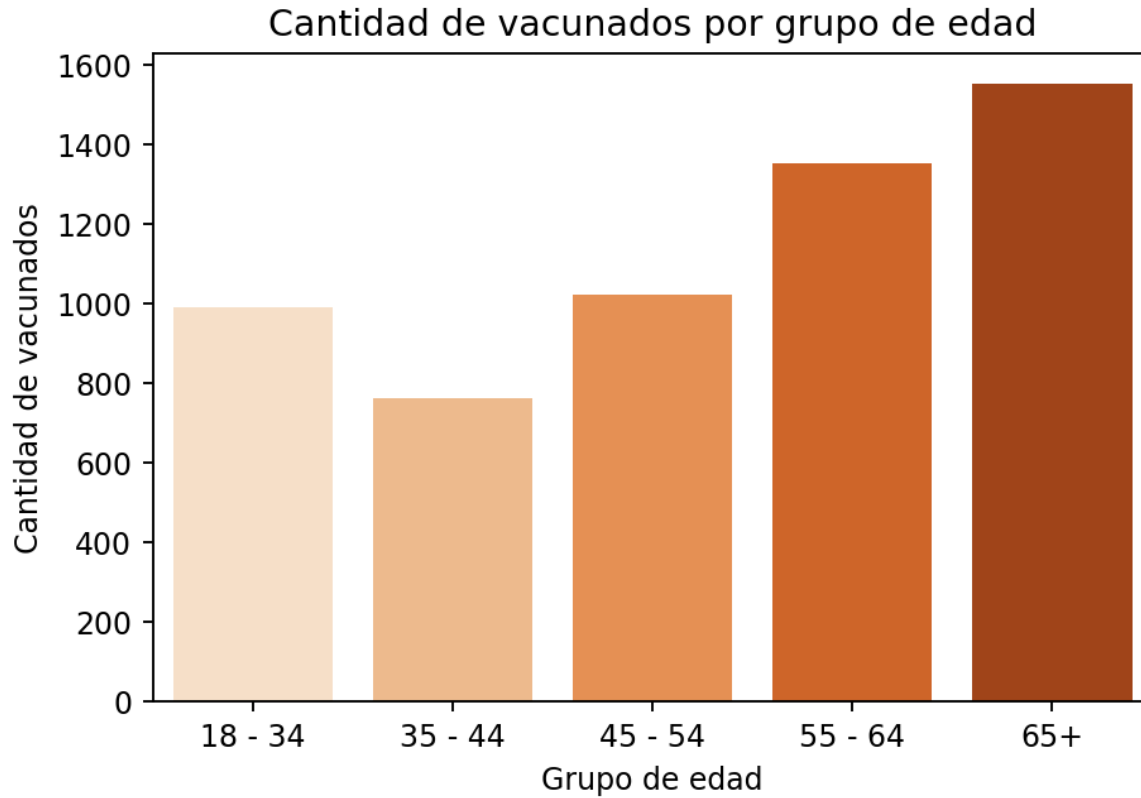
Visualización 5: HeatmapPlot – Predict H1N1 and Seasonal Flu Vaccines –

Cantidad de vacunados según relación conocimiento/preocupación h1n1



<https://github.com/MauRodriguez/tp1.datos.2021.2/blob/master/visu/visu-vaccines.ipynb>

Visualización 6: BarPlot – Predict H1N1 and Seasonal Flu Vaccines –



<https://github.com/MauRodriguez/tp1.datos.2021.2/blob/master/visu/visu-vaccines-2.ipynb>

Visualización 7: Mejores 10 winratios en la historia de F1

La idea es mostrar cuales son los 10 mejores pilotos que mejor se desempeñaron hasta el día de hoy en la F1 tomando en cuenta cantidad de salidas (starts) comparado con cantidad de victorias (wins). Se muestra también, entre otras cosas, los autos más icónicos de cada corredor, lo que denota las distintas épocas donde se desempeñan, y los números de WINS/STARTS lo que denota que algunos corredores tuvieron mucho mejor desempeño sin tener tanta experiencia corriendo.

FUENTE: <https://www.kaggle.com/rohanrao/formula-1-world-championship-1950-2020>

Pequeño análisis exploratorio:

<https://github.com/MauRodriguez/tp1.datos.2021.2/blob/master/visu/F1.ipynb>

F1 HISTORIC BEST WIN RATIO (WINS/STARTS)

J.M. FANGIO
41.3% WR
24W | 58S



L. HAMILTON
35.6% WR
98W | 275S



M. SCHUMACHER
29.5% WR
91W | 308S



A. SENNA
25.3% WR
41W | 162S



S. MOSS
21.9% WR
16W | 73S




A. ASCARI
36.1% WR
13W | 36S


J. CLARK
34.2% WR
25W | 73S


J. STEWART
27.0% WR
27W | 100S


A. PROST
25.2% WR
51W | 202S


S. VETTEL
19.8% WR
53W | 267S