

Manejo de datos con pandas y K Means

BRANDON, SALVADOR Y ROY





En esta actividad se espera aprender a utilizar pandas y KMeans entendiendo como funciona el machine learning y empleando los conocimientos para procesar un conjunto de datos.



Work experience



GENERAR PROYECTO DE GITHUB

USO DE GIT BASH

GENERAR DATOS ESTADÍSTICOS Y GRÁFICAS

Pandas y Numpy

IDENTIFICAR PATRONES

Analizar los gráficos generados

GENERAR K-MEANS

librería sklearn.cluster

ANÁLISIS DE RESULTADOS

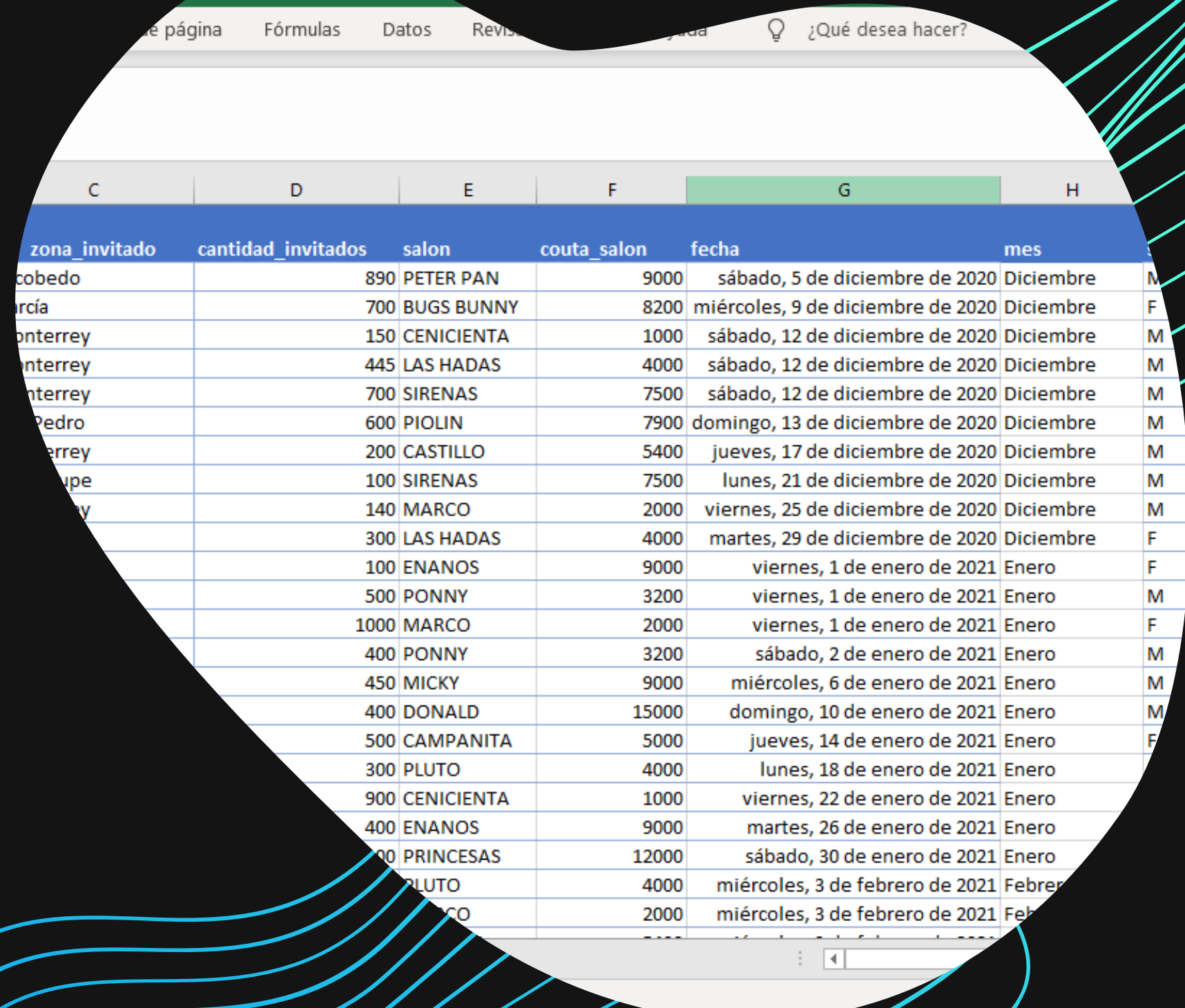
Observar las respuestas de los gráficos y análisis

Repositorio Git Hub



Proyecto

Se tomó un archivo excel con datos sobre fiestas. Antes que nada se modificaron los heads del excel para un mejor manejo:

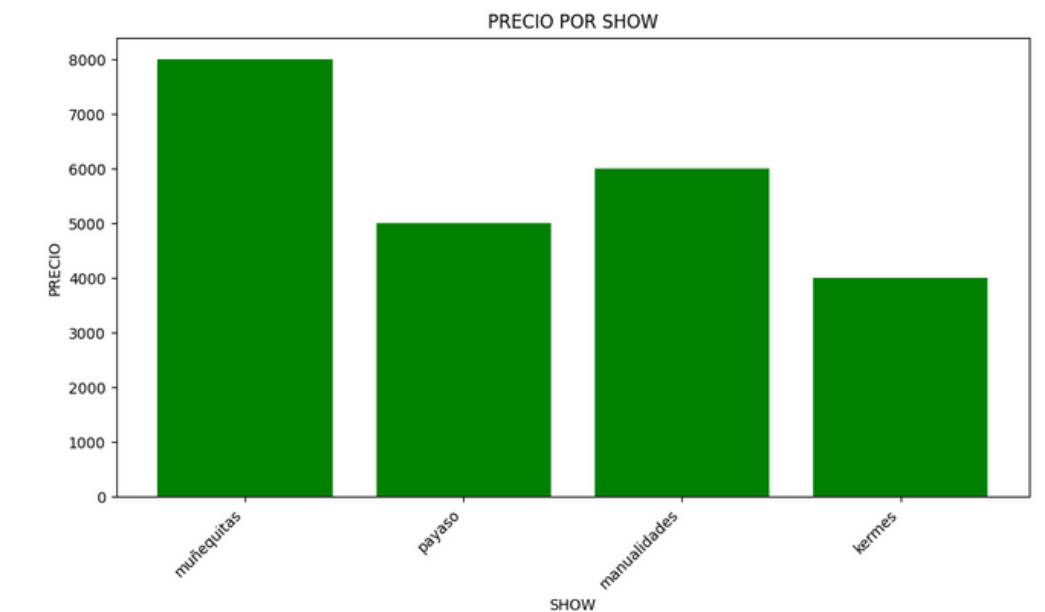
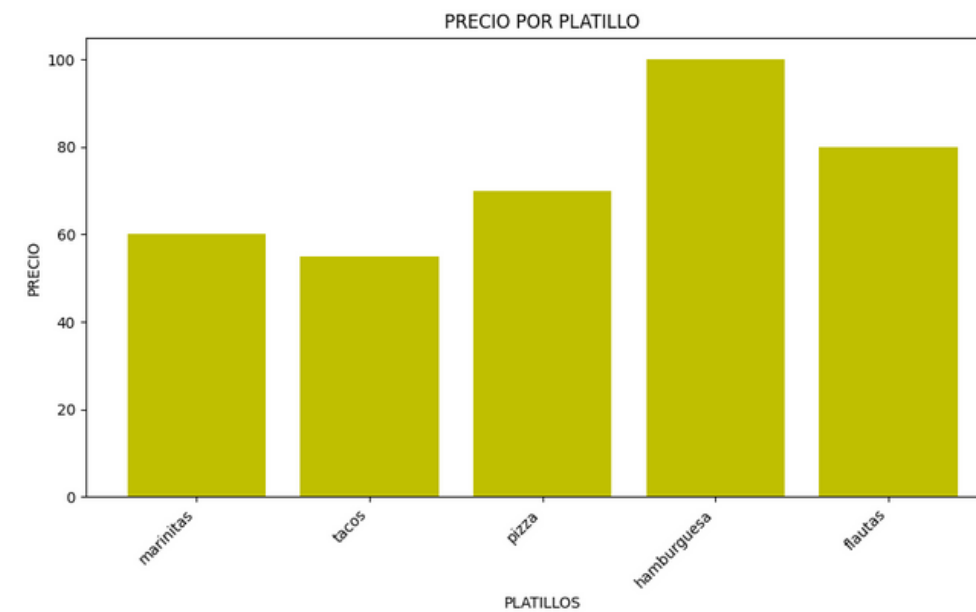
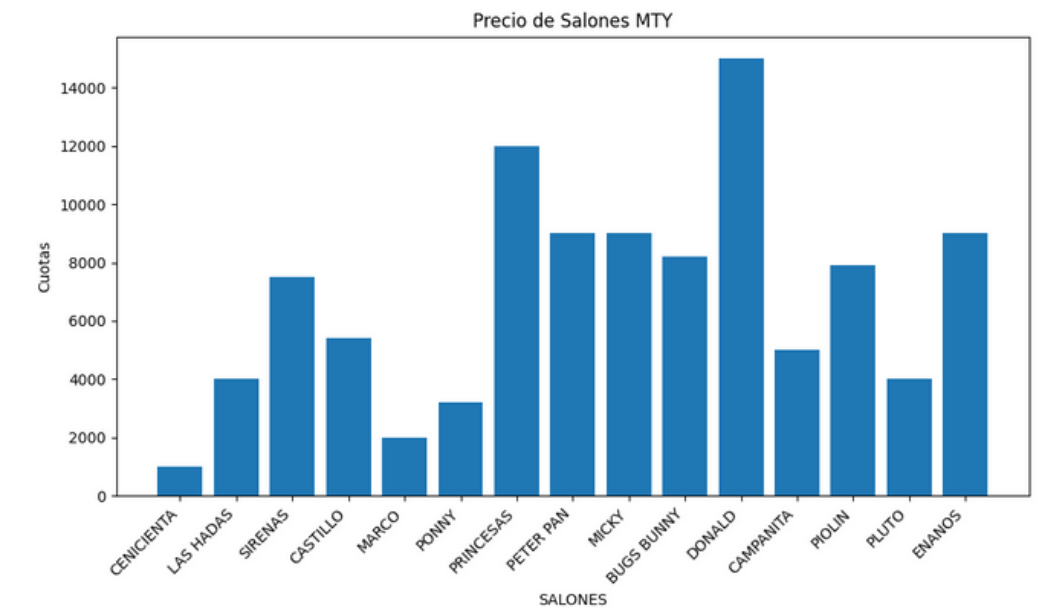
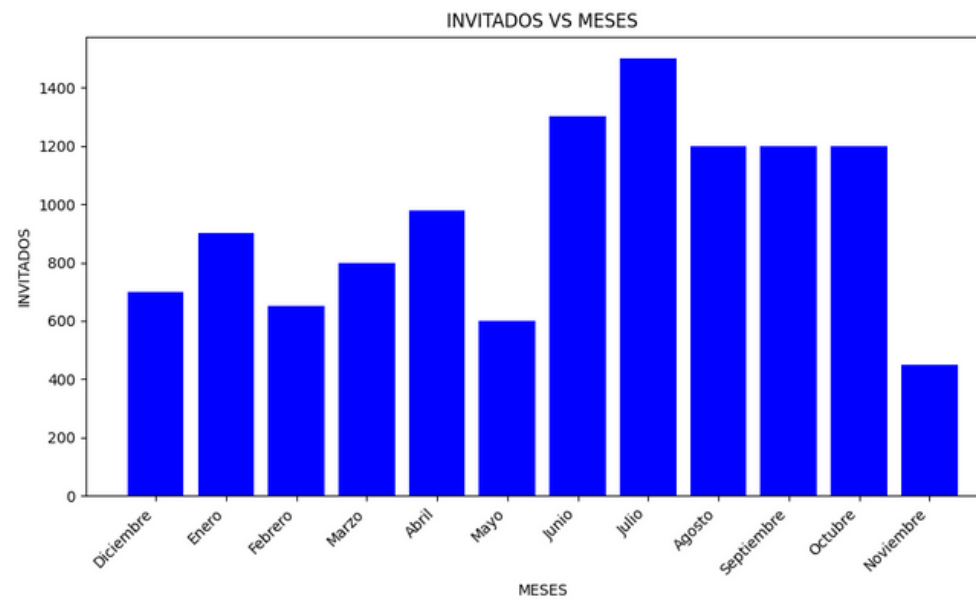


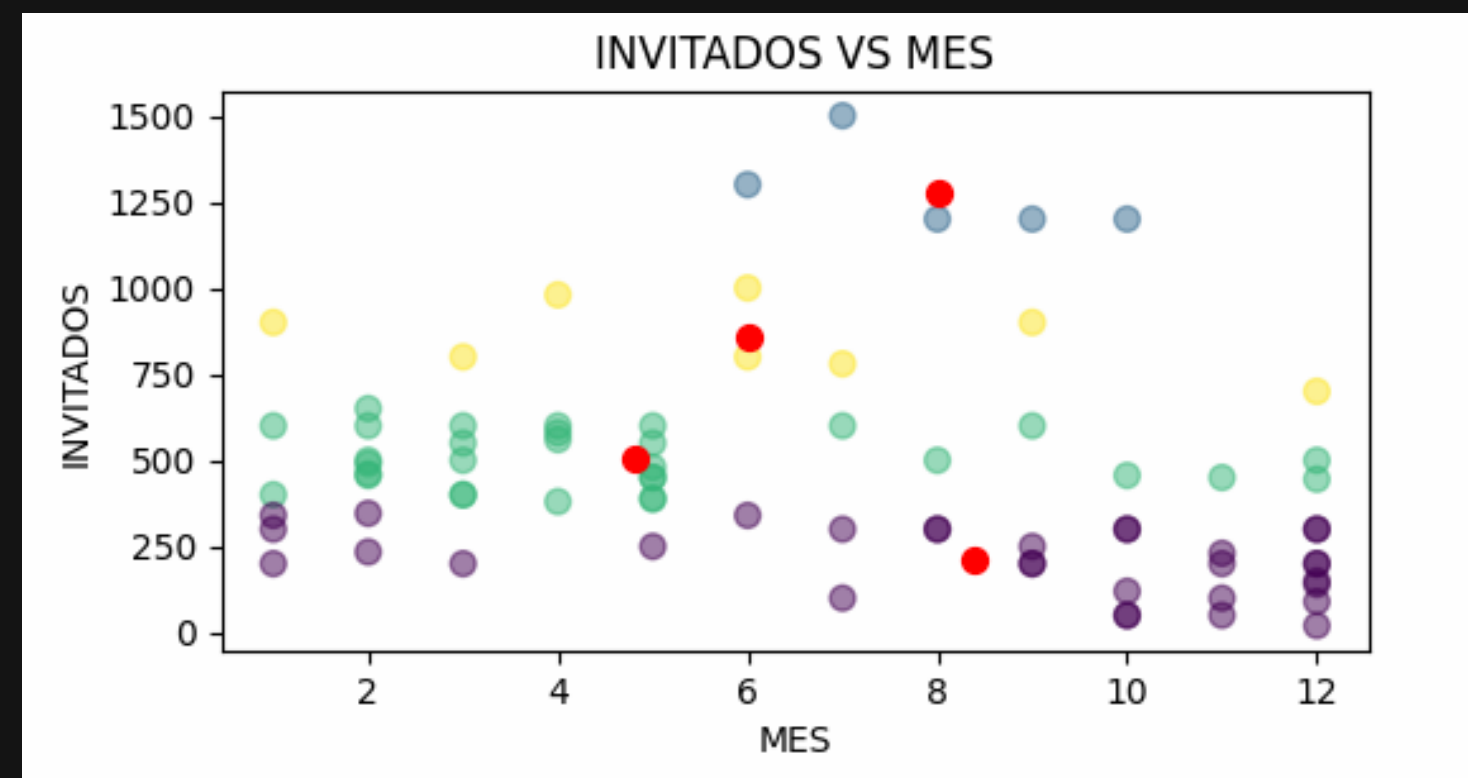
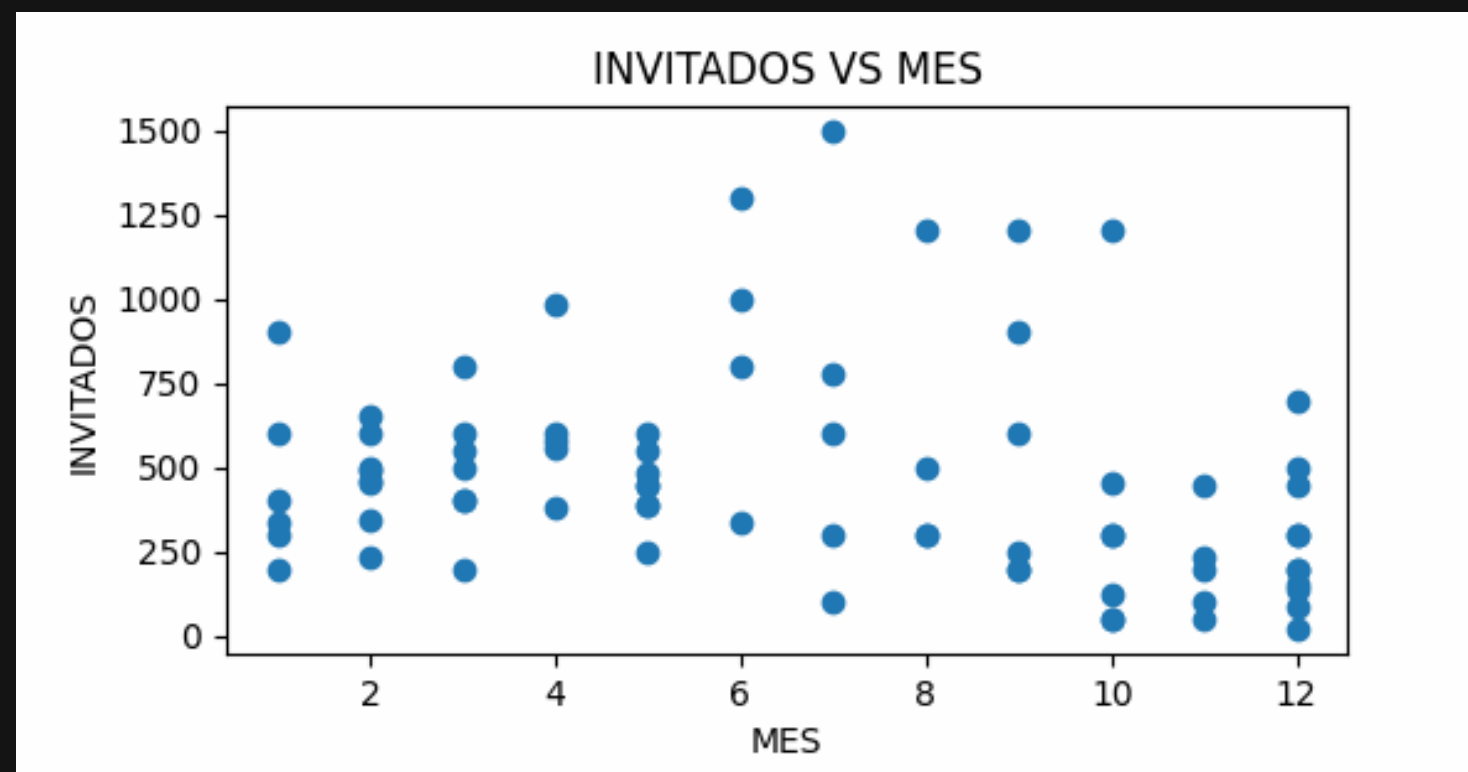
zona_invitado	cantidad_invitados	salon	couta_salon	fecha	mes
cobedo	890	PETER PAN	9000	sábado, 5 de diciembre de 2020	Diciembre
arcía	700	BUGS BUNNY	8200	miércoles, 9 de diciembre de 2020	Diciembre
onterrey	150	CENICIENTA	1000	sábado, 12 de diciembre de 2020	Diciembre
onterrey	445	LAS HADAS	4000	sábado, 12 de diciembre de 2020	Diciembre
onterrey	700	SIRENAS	7500	sábado, 12 de diciembre de 2020	Diciembre
Pedro	600	PIOLIN	7900	domingo, 13 de diciembre de 2020	Diciembre
errey	200	CASTILLO	5400	jueves, 17 de diciembre de 2020	Diciembre
upe	100	SIRENAS	7500	lunes, 21 de diciembre de 2020	Diciembre
y	140	MARCO	2000	viernes, 25 de diciembre de 2020	Diciembre
	300	LAS HADAS	4000	martes, 29 de diciembre de 2020	Diciembre
	100	ENANOS	9000	viernes, 1 de enero de 2021	Enero
	500	PONNY	3200	viernes, 1 de enero de 2021	Enero
	1000	MARCO	2000	viernes, 1 de enero de 2021	Enero
	400	PONNY	3200	sábado, 2 de enero de 2021	Enero
	450	MICKY	9000	miércoles, 6 de enero de 2021	Enero
	400	DONALD	15000	domingo, 10 de enero de 2021	Enero
	500	CAMPANITA	5000	jueves, 14 de enero de 2021	Enero
	300	PLUTO	4000	lunes, 18 de enero de 2021	Enero
	900	CENICIENTA	1000	viernes, 22 de enero de 2021	Enero
	400	ENANOS	9000	martes, 26 de enero de 2021	Enero
	1000	PRINCESAS	12000	sábado, 30 de enero de 2021	Enero
	4000	PLUTO	4000	miércoles, 3 de febrero de 2021	Febrero
	2000	MARCO	2000	miércoles, 3 de febrero de 2021	Febrero

Parte 1

Se utilizó la librería pandas para hacer gráficas para observar los datos:

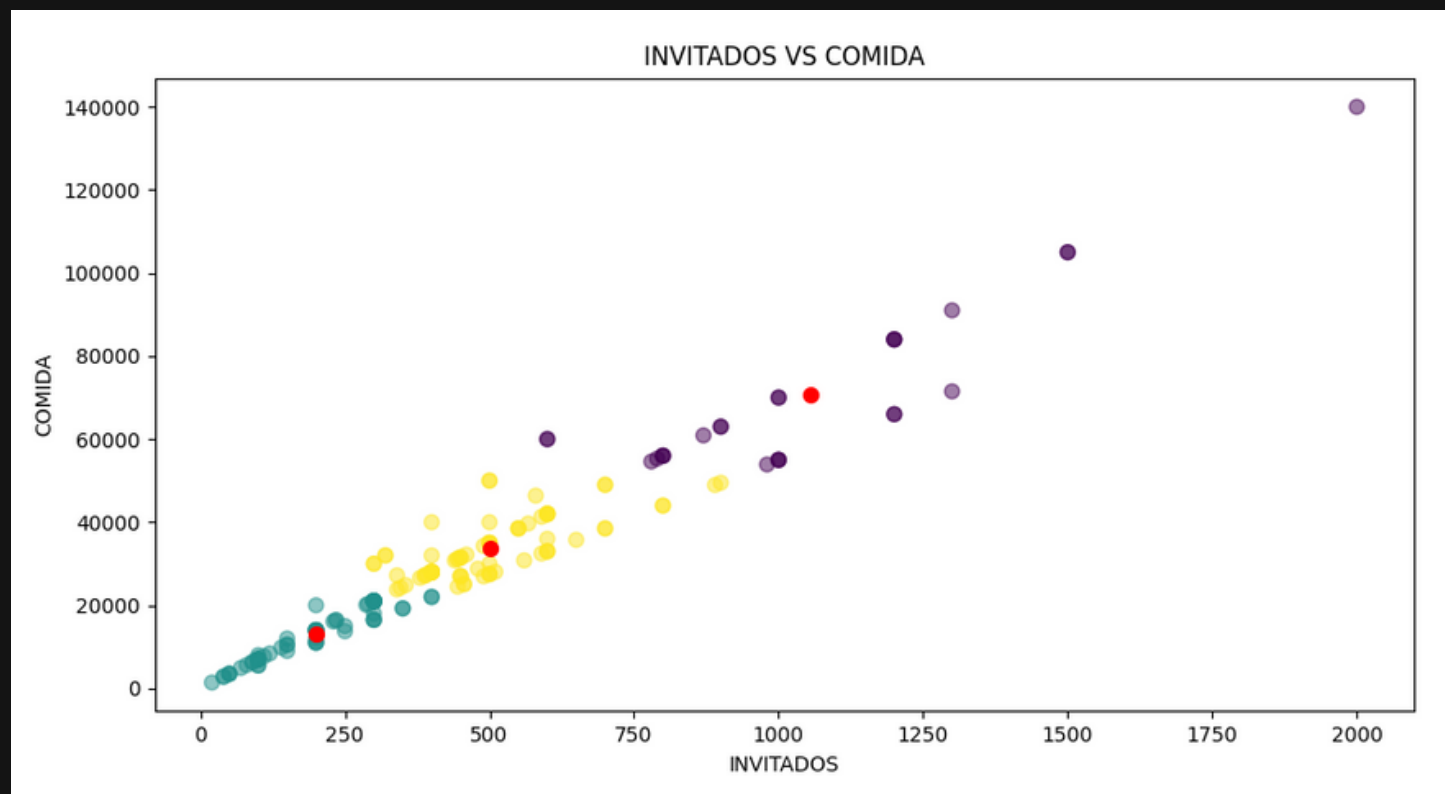
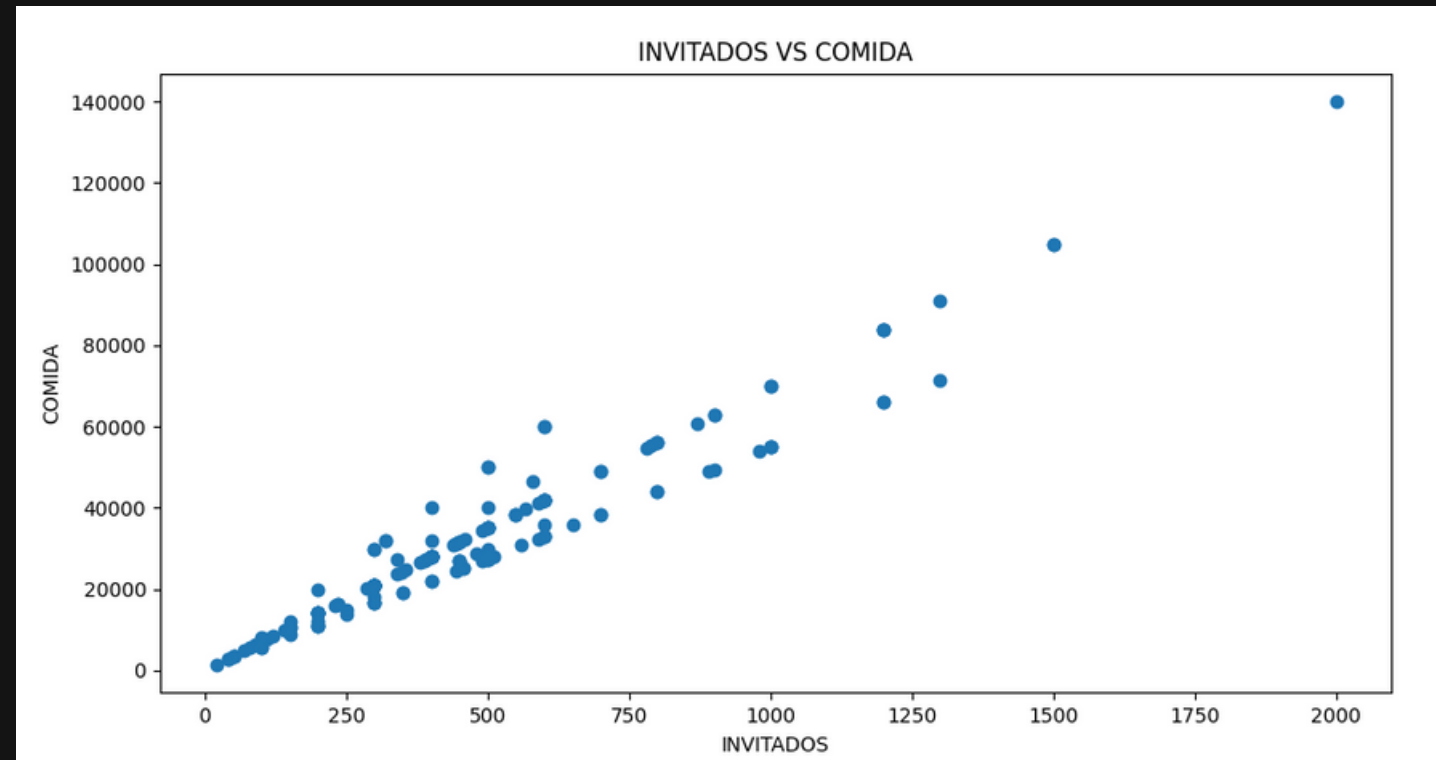
- Invitados vs meses
- Precios de salones
- Precio por platillo
- Precio por show





PARTE DE BRANDON

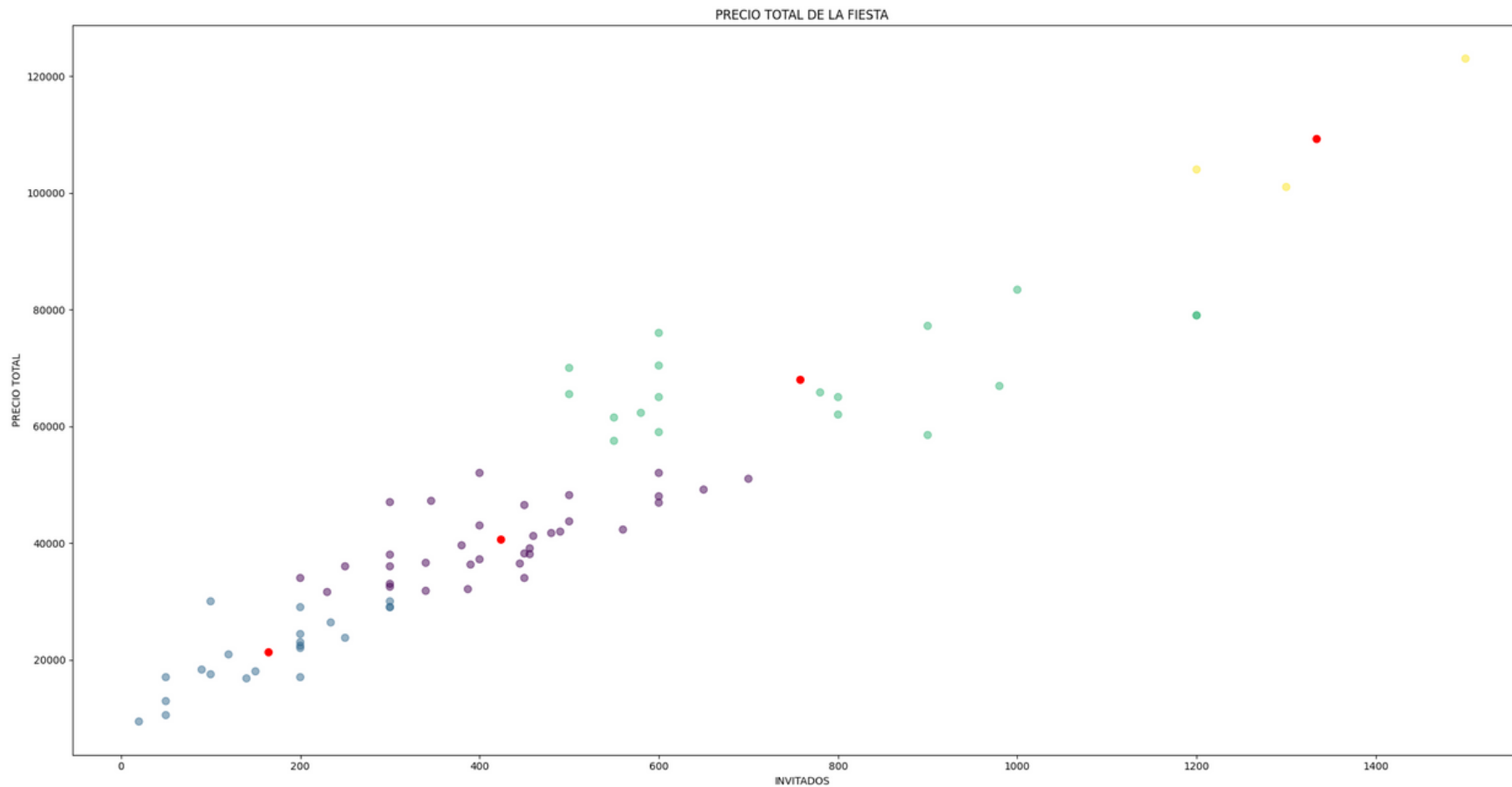
Con ayuda de kmeans se obtuvieron los clústers de invitados de Monterrey con los meses. Para este caso se eligieron 4 clústers porque se buscaba segmentar el rango en el que se podría realizar una fiesta, en el que con 4 se obtiene una buena segmentación (si se le agregan más se segmenta demasiado)



PARTE DE SALVADOR

Con ayuda de kmeans se obtuvieron los clusters de invitados contra precio de alimentos en todas las regiones.

En este caso se decidió hacer una segmentación de 3 clusters, esto ya que se busca analizar 3 secciones de costo total de comida con respecto al num de invitados.



PARTE ROY

Se crearon los clusters de precios totales contra número de invitados tomando en cuenta sólo las fiestas que se llevaron a cabo en Monterrey. Se usaron 4 clusters para determinar las fiestas pequeñas, medianas, grandes y las que se salen del rango normal: las enormes.