

Maestría en Ciencia de Datos

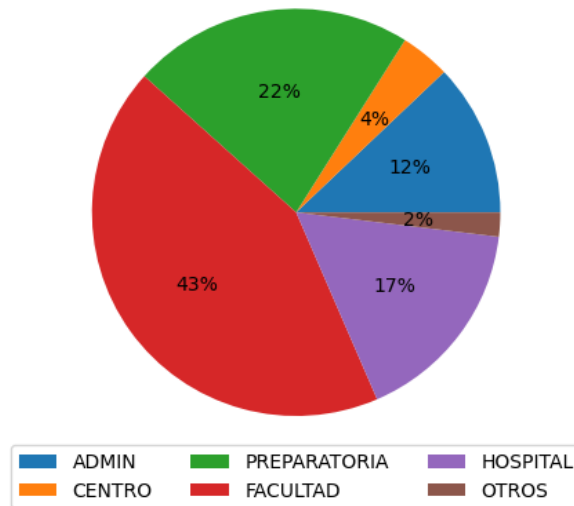
Aprendizaje Automático

Alvaro Pequeño Mondragón
1726520

Reporte #1

30 de junio de 2024

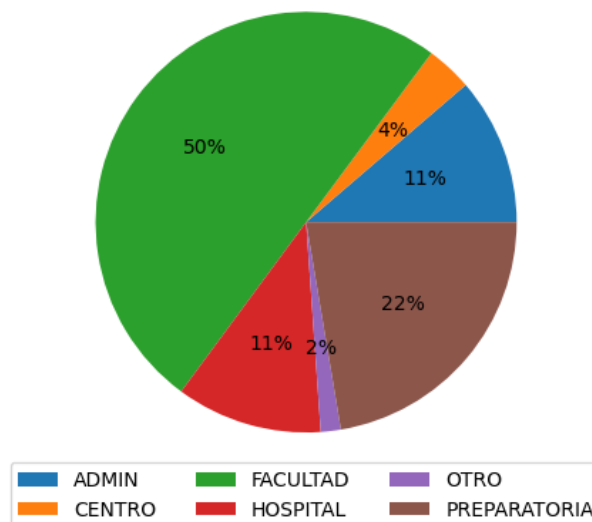
Porcentaje de Pagos por Tipos de Institución



La imagen anterior muestra la cantidad de pagos que se hacen por tipo de dependencia y lo que me parece curioso es que el hospital tiene una gran cantidad de pagos aun siendo 1 sola entidad, superando a otras dependencias que se conforman de mas entidades y está casi a la par de las preparatorias.

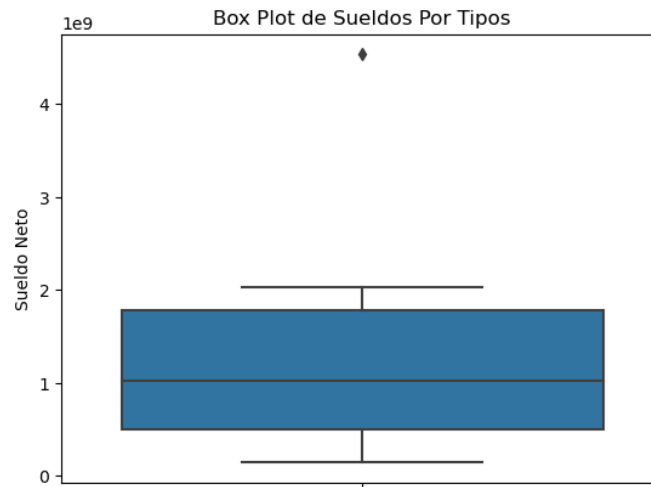
Después de lo anterior decidí ver una gráfica similar pero que tuviera el total de dinero gastado por tipo de dependencias:

Porcentaje de Dinero Pagado por Tipos de Institución



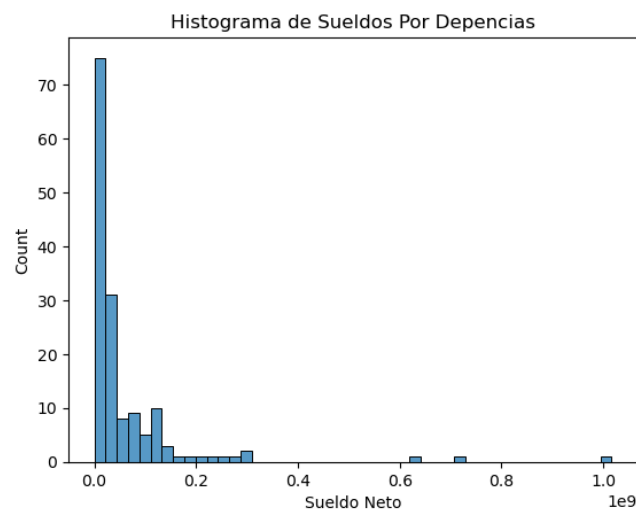
En la imagen anterior se observa como las facultades son las que mas gastan siendo el 50% de todos los gastos en salarios de la universidad. Al ver la cantidad de dinero gastada por el hospital y compararla con la cantidad de pagos es curioso que, aunque en cantidad de pagos el hospital tiene un 5% más que los administrativos en cuestión de cantidad de dinero ambos tienen el 11% por lo que es posible que los sueldos en la parte administrativa sean un poco más altos que en el hospital.

Solo por no dejar decidí hacer un boxplot con la cantidad de sueldos pagada por dependencias:



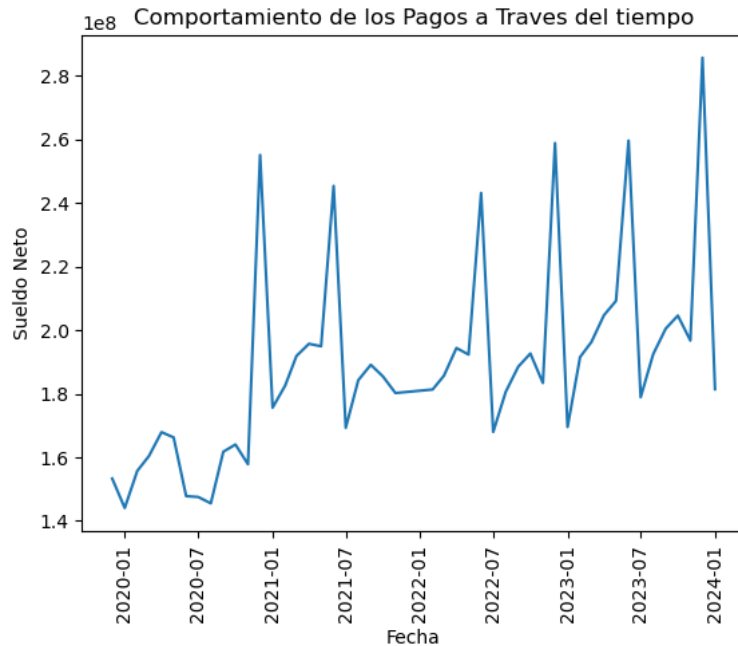
Si recordamos en la grafica de pastel las facultades representan el 50% del dinero gastado en sueldos y esto se ve reflejado en el boxplot como el punto atípico que se muestra en la parte superior de la gráfica.

Después de ver lo anterior me interesó saber como es que se distribuyen los salarios de todas las dependencias:

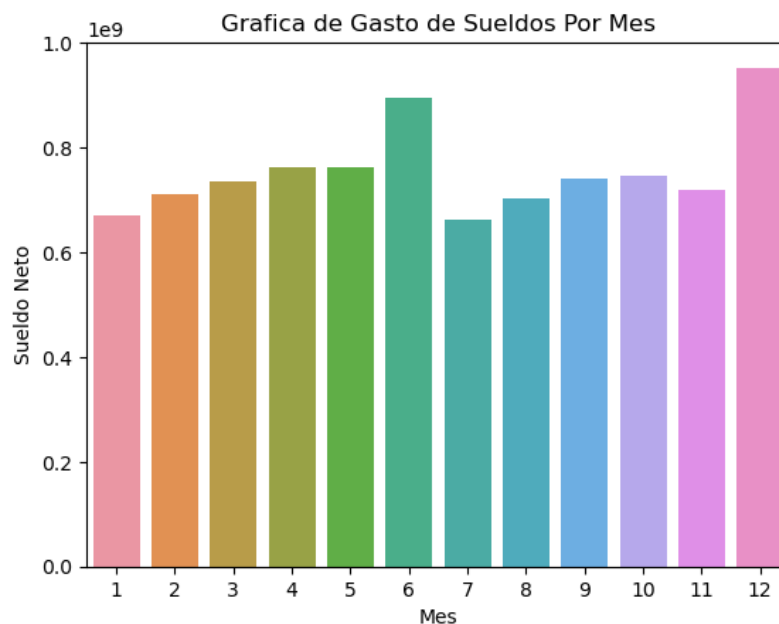


Al observar la información de la grafica se puede notar que la gran mayoría de las dependencias tienen su gasto total en sueldos en una cantidad menor a 200,000,000 de pesos y después de esta cantidad son contadas las dependencias que llegan a tener un gasto mayor a eso.

Después quise replicar una grafica que vimos en clase que es una serie de tiempo:

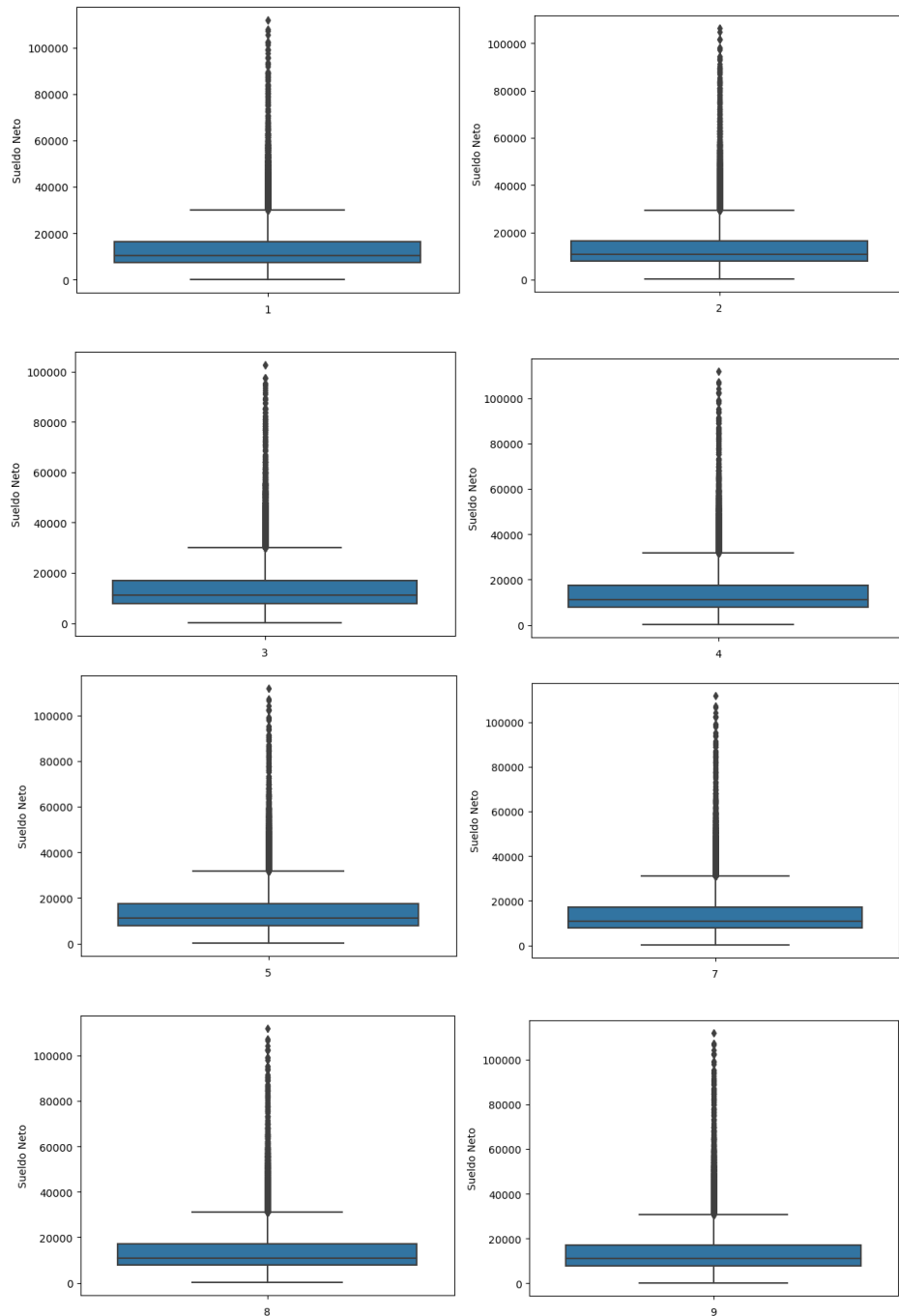


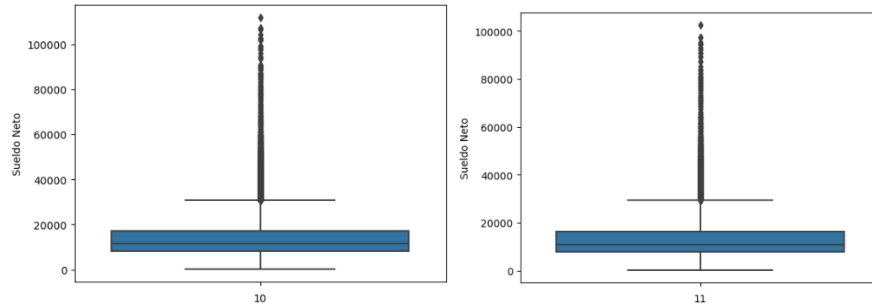
Al verla recordé como se discutió en clase que los picos representaban aquellos meses donde se entregan prestaciones como el aguinaldo o diversos bonos, por lo que después de eso decidí hacer una gráfica de barras por mes parera ver como se ven los meses en cuestión de sueldos pagados:



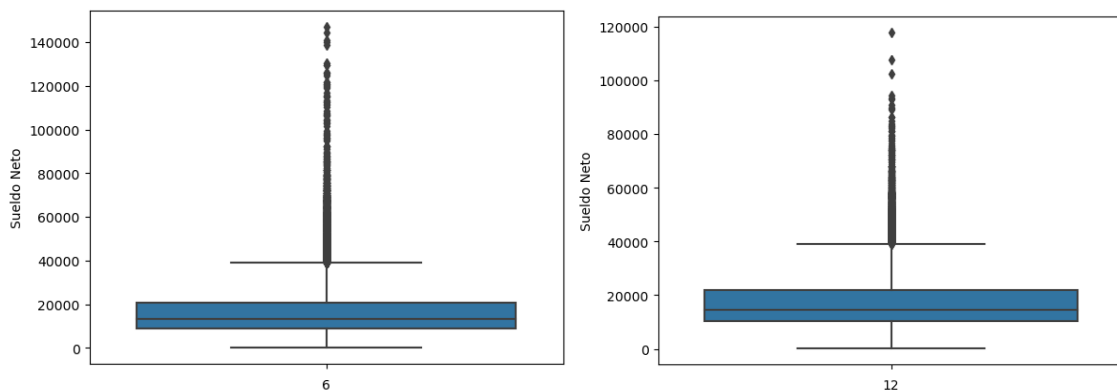
En la grafica anterior podemos ver como los 2 meses que mayor cantidad de sueldo tienen son diciembre y junio por lo que podemos decir que en esos 2 meses la universidad paga las prestaciones y bonos. Además, podemos ver como en el siguiente mes de junio y diciembre cae el sueldo que se paga y luego empieza a subir lentamente hasta llegar a un nuevo pico, este comportamiento se puede ver de una mejor manera en la serie de tiempo.

Una vez visto lo anterior decidí realizar los ciclos para generar imágenes y esto es algo que nunca había hecho por lo que me sorprendió la facilidad y rapidez con la que se pueden generar las diferentes imágenes, y solo para reforzar lo que se vio en las graficas anteriores en cuanto a los pagos por mes así se ven los boxplots generados mediante un ciclo de los sueldos por mes:





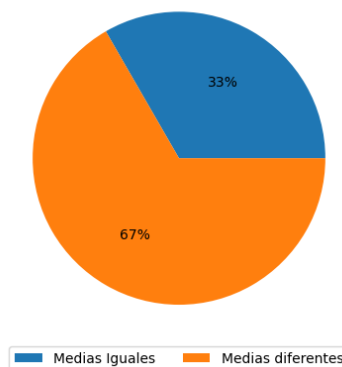
En las graficas anteriores se muestra los meses donde no hay prestaciones o bonos, y se puede ver que la escala llega hasta 100,000 mientras que si observamos las graficas de los meses donde se pagan las prestaciones o bonos la escala del boxplot llega a 140,000 mostrando el incremento en pagos:



Después de esto la tarea se continuó con las pruebas estadísticas, por lo que quise ver de forma grafica alguno de los resultados que se obtuvieron en esa parte. Primero hice una prueba entre tipo de dependencias y resulto que ninguna dependencia tenía sus medias iguales, por lo que tenía la duda de como se vería algo que tuviera sus medias iguales, por lo que decidí hacer la prueba a entidades del mismo tipo, en este caso las preparatorias.

Al hacer la prueba de Kruskal-Wallis salió que no todas las medias iguales, pero al hacer la prueba de tukey y probar por pares salieron varios que si tenían sus medias iguales y a partir de eso se obtuvo la siguiente gráfica:

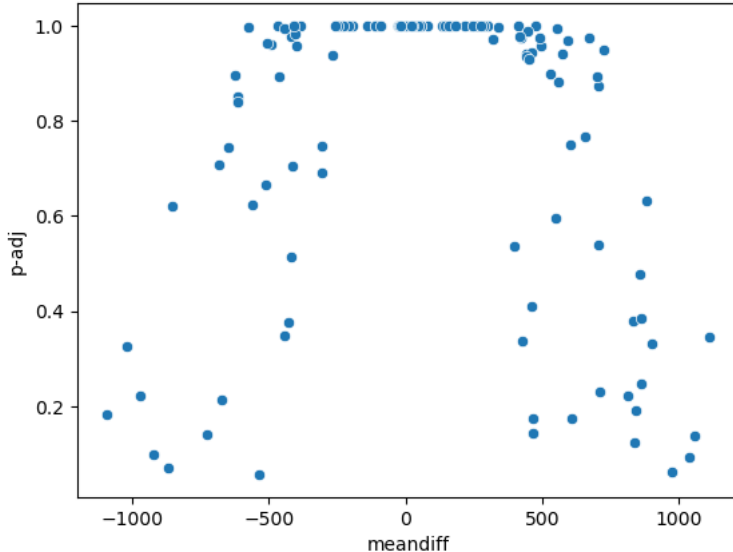
Prueba de Tukey: Pares con media igual y pares con media diferente



En la grafica se puede observar que de todos los pares posibles el 33% tienen medias iguales por lo que significa que trabajar en una u otra prepa debería tener una cantidad similar de pago.

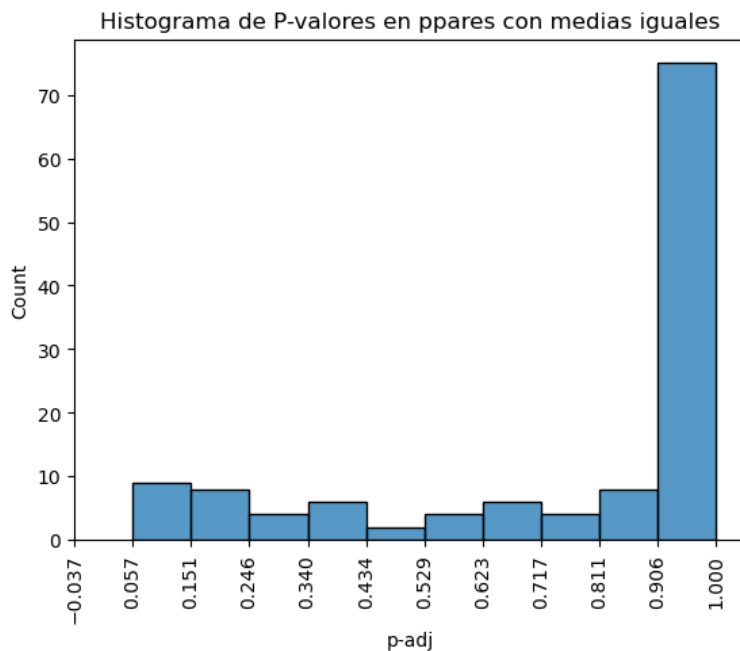
Después quise ver como es que se comporta los p-valores conforme va cambiando la diferencia de medias en aquellos pares donde la prueba retorno que tienen medias iguales:

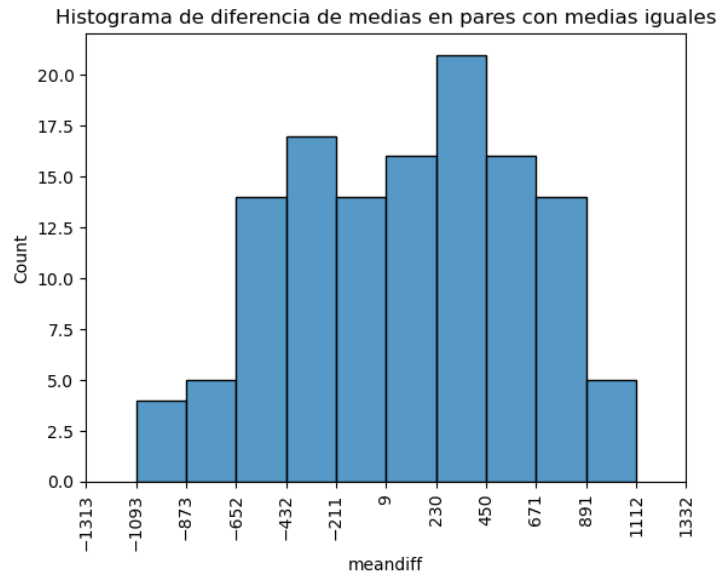
Prueba de Tukey: Comportamiento de P-valor conforme cambia la diferencia de medias



Fue interesante ver un comportamiento simétrico, y que aproximadamente en una diferencia de 250 negativa o positiva el p-valor es prácticamente 1. También podemos ver como se considera medias iguales cuando la diferencia es de 1000 o menos.

Por último, tenía interés en ver cómo era la distribución de los p-valor y también de las diferencias de medias en aquellos pares que tienen medias iguales:





Antes de generar las graficas yo pensé que la distribución de los p-valores sería idéntica al de las diferencias de medias, pero al ver las graficas generadas me encontré con que estaba equivocado las graficas son diferentes y no se parecen en lo más mínimo.