**INFORME**

1. **Análisis de Penguins**

Descripción de variables:

- Species: especie de pinguino (variable categorica)

- Island: isla donde se observo al pinguinito (variable categorica)

- bill\_length\_mm: largo del pico en milímetros (variable numérica continua)

- bill\_depth\_mm: ancho del pico en milímetros (variable numérica continua)

- flipper\_length\_mm: largo de las alas en milímetros (variable numérica continua)

- body\_mass\_g: masa del pinguino en gramos (variable numérica continua)

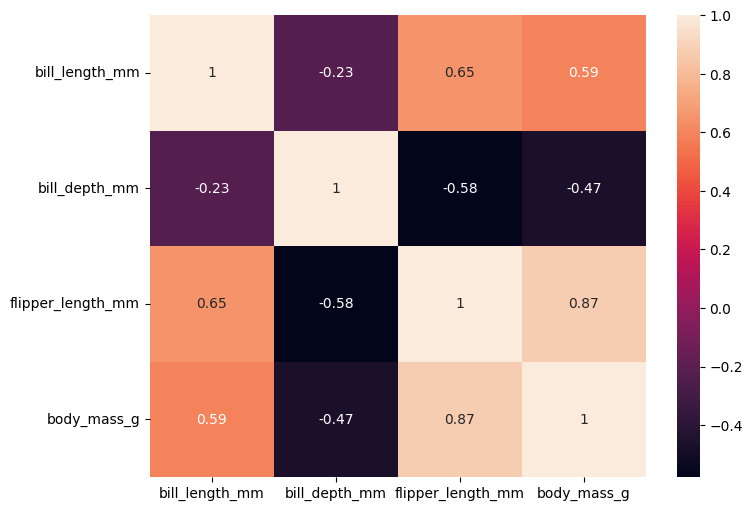
- Sex: genero del pinguino (variable categorica)

El dataset, originalmente, contaba con 344 observaciones y 7 variables, pero por la presencia de varios valores nulos, se redujo a 333 observaciones. Detalle de datos faltantes:



Estadísticos del Dataset:

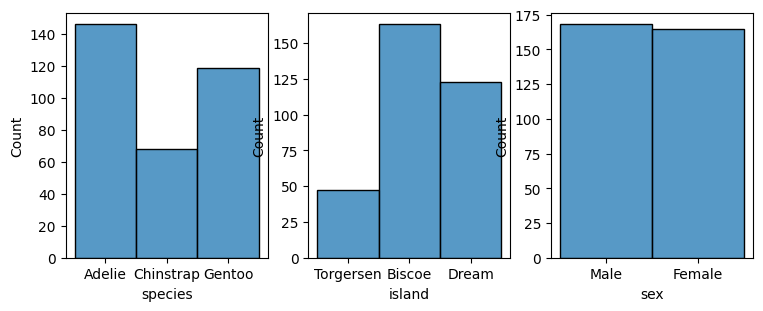




*Gráfica 1.1: Covarianzas*

Análisis Exploratorio de Datos (AED)

Hallazgos:

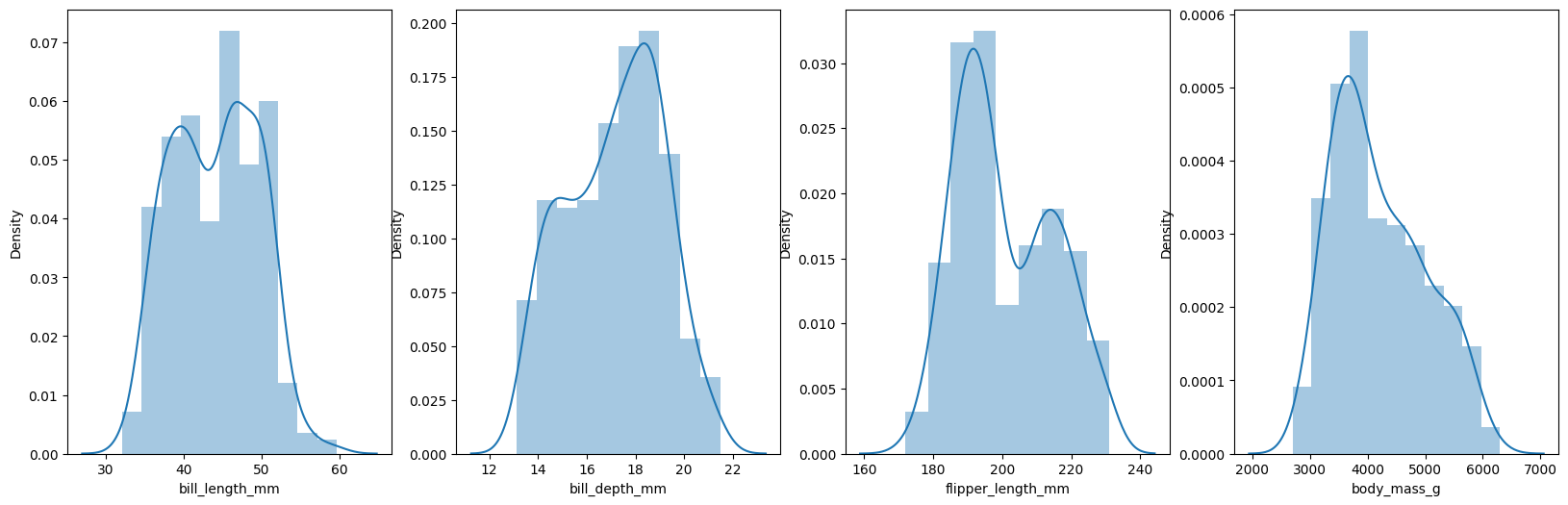


*Gráfica 1.2: Grafica de barras de variables categóricas*

- La distribución observada de machos y hembras en la observación es muy similar, indicando que son casi el mismo número de machos y hembras

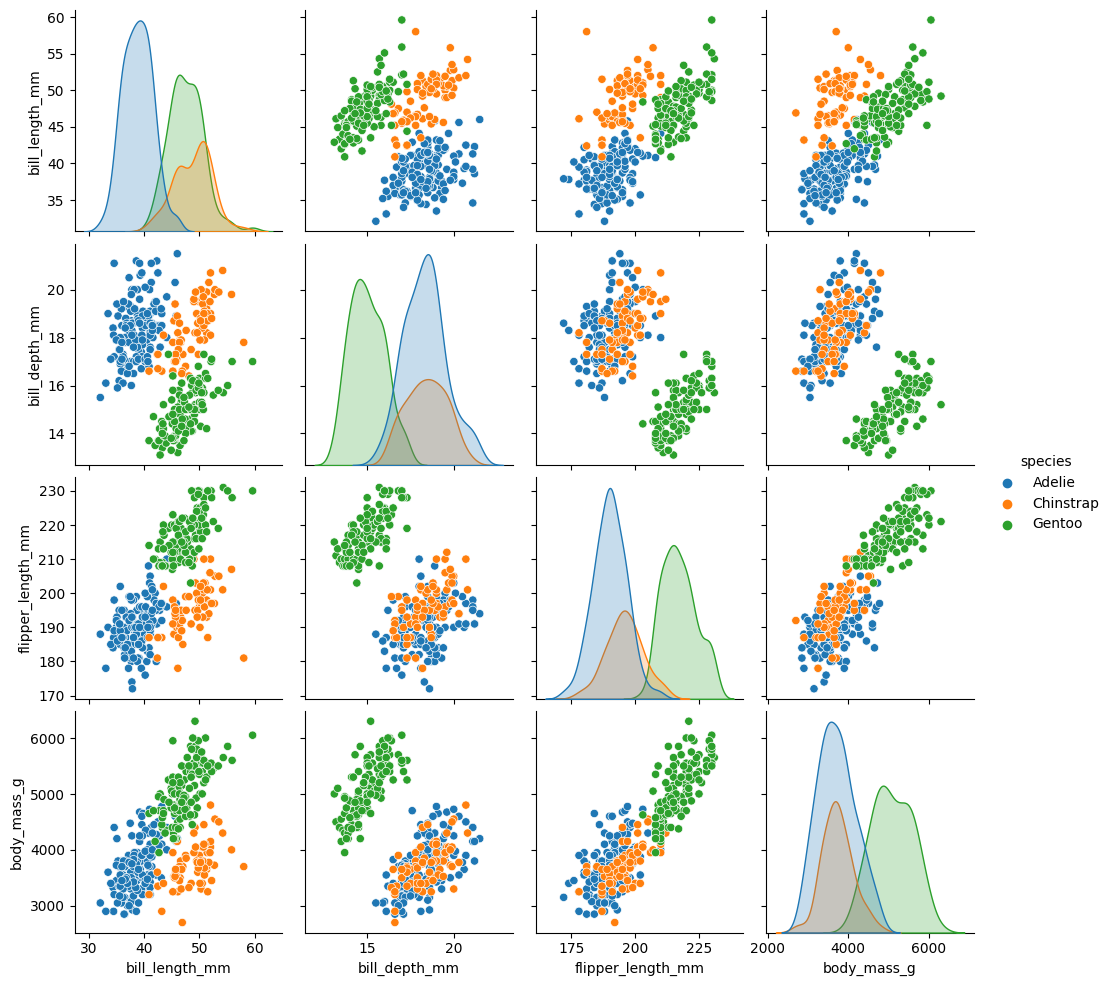
- Hubo más observaciones en Biscoe y Dream que en Torgersen

- En mínimo 2 islas hay presencia de por lo menos 2 especies de pingüinos, ya que Biscoe tiene más pingüinos de la cantidad reportada para el género Adelie. Caso contrario, en el que en cada isla solo hubiera una especie, tendría que haber similitudes más marcadas entre los histogramas de islas y especies.



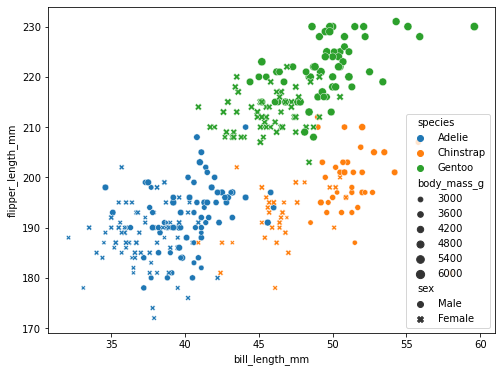
*Gráfica 1.3: Distribuciones variables numéricas.*

En el caso de la primera gráfica, se puede apreciar una leve bimodalidad, mientras que las otras tienen caídas más marcadas y tendencias no tan delimitadas. Estas tendencias las separamos en especie para ver mejor cómo se comporta por especie.



*Gráfica 1.4: Scatter plot entre variables y diagramas de densidad de cada variables numéricas. Separado por especie.*

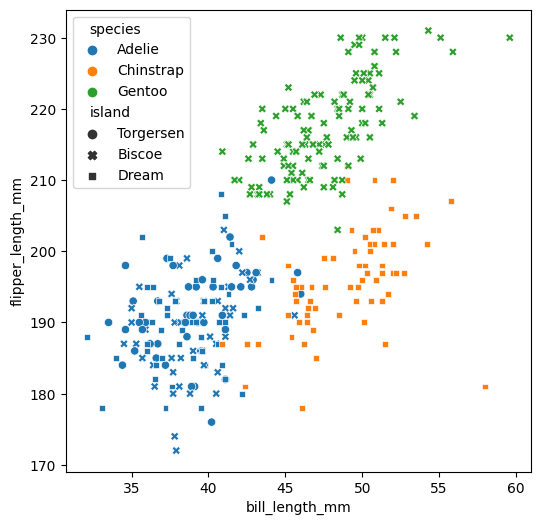
* En la gráfica 1.1 vimos una bimodalidad ya que Adelie y Gentoo, que son las especies predominantes, tenían una diferencia en la relación largo pico y masa corporal.
* Adelie y Chinstrap tienen distribuciones corporales similares ya que son especies muy parecidas en forma. En cambio, Gentoo tiene diferencias más marcadas y por eso su distribución si varía considerablemente contra las otras dos especies.



*Gráfica 1.5: Scatter plot entre flipper length y bill length. Separada por color para especie, por tamaño para masa corporal y por simbolo para la variable de sexo.*

Los machos tienden a ser en todas las especies más grandes y pesados. Aunque sí varían entre especies por diferencias mencionadas previamente, si se mantiene que el macho es más pesado y grande que la hembra.

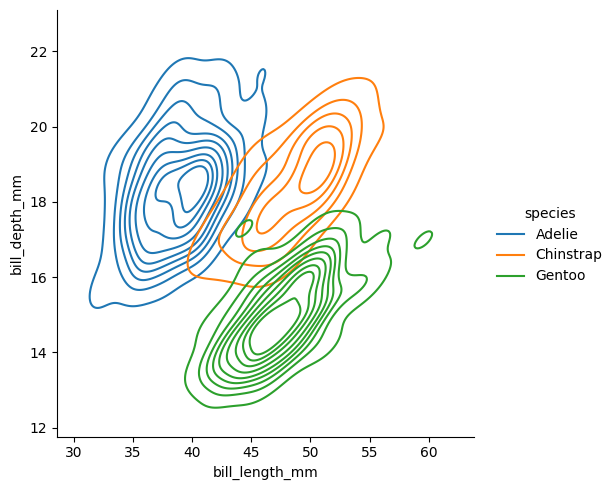
La especie Adelie tiende a ser la más pequeña, porque tanto machos como hembras tienen masas corporales más pequeñas que las otras dos especies.



*Gráfica 1.6: Scatter plot entre flipper length y bill length. Separada por color para especie y por simbolo para la variable de isla.*

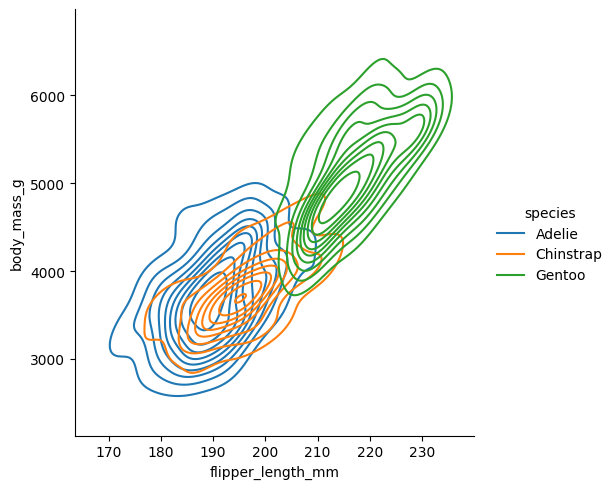
- Chinstrap solo tiene observaciones en la isla Dream y Gentoo solo tiene observaciones en Biscoe

- Adelie tiene presencia en las tres islas, confirmando el supuesto de arriba, pero solo se encontró esta especie en la isla Torgersen



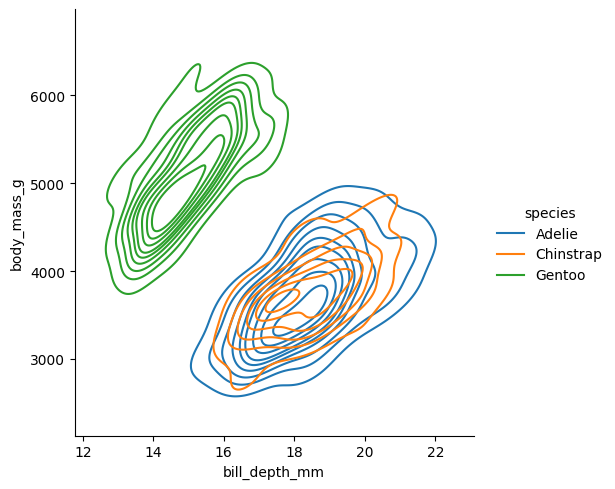
*Gráfica 1.7: Diagrama de calor de flipper length y bill length. Separado por especie.*

Los tamaños de los picos varían por cada especie y si hay diferencias marcadas.



*Gráfica 1.8: Diagrama de calor de flipper length y body mass. Separado por especie.*

La distribución corporal (largo pico y masa) son parecidas entre Adelie y Chinstrap por razones ya descritas previamente, mientras que Gentoo si presenta un cambio más marcado.



*Gráfica 1.8: Diagrama de calor de bill depth y body mass. Separado por especie.*

La relación entre el ancho del pico y la masa es idéntica entre Adelie y Chinstrap, con una variación en la media de ambos. En este caso, no hay colindancia entre Gentoo y las otras dos especies.

Conclusiones:

* Existen diferencias muy marcadas entre la especie Gentoo y las otras dos especies observadas, Adelie y Chinstrap. Principalmente de las proporciones de su cuerpo, se observa que la razón entre masa corporal y largo del pico y largo del ala es casi la misma para las últimas dos especies y que quienes presentan una proporción distinta es Gentoo. (Gráfica 1.4, 1.7 y 1.8)
* La especie de Adelie es la más pequeña de las observadas (gráfica 1.5), pero al mismo tiempo es la que se encuentra en más islas, está presente en las tres islas observadas. Por otra parte Gentoo y Chinstrap, solo se encontraban en una isla distinta cada especie.
* Las hembras de cada especie siempre tienden a ser más pequeñas que los machos, esto se debe a que presentan una masa corporal menor y un tamaño de ala y de pico más pequeño. (Gráfica 1.5)

1. **Análisis de Tips**

Descripción de variables:

- Total\_bill: total de la cuenta (variable numérica continua)

- Tip: cantidad de propina que recibió el mesero (variable numérica continua)

- Sex: sexo del que pagó la cuenta (variable categorica)

- Smoker: si alguno de los de la mesa fumaba o no (variable categórica)

- Day: dia de la semana (variable categórica)

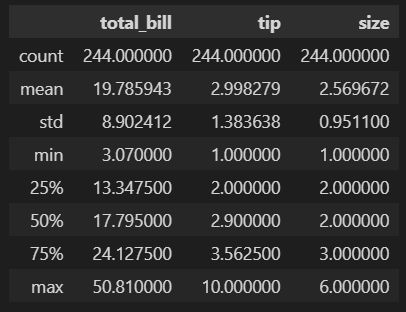
- Time: time of the day (variable categórica)

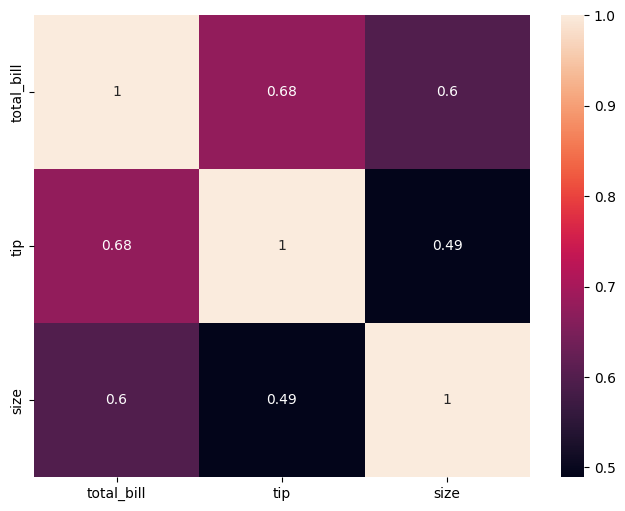
- Size: cantidad de personas en la mesa (variable numérica discreta)

El dataset cuenta con 244 observaciones y 7 variables. No se encontraron valores nulos por lo que se trabajó con la totalidad del dataset



Estadísticos del Dataset

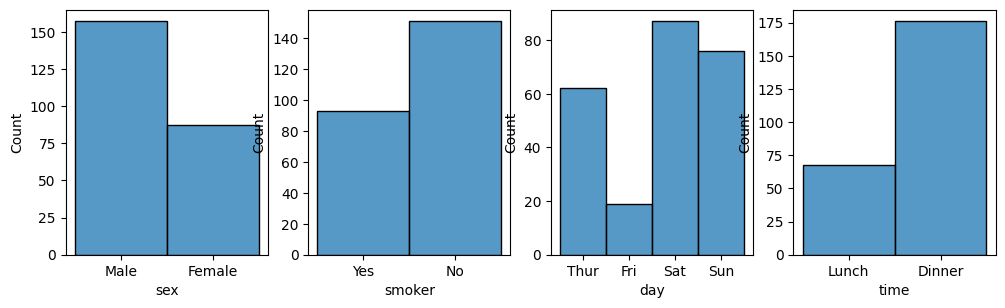




*Gráfica 2.1: Covarianzas*

Análisis Exploratorio de Datos (AED)

Hallazgos:



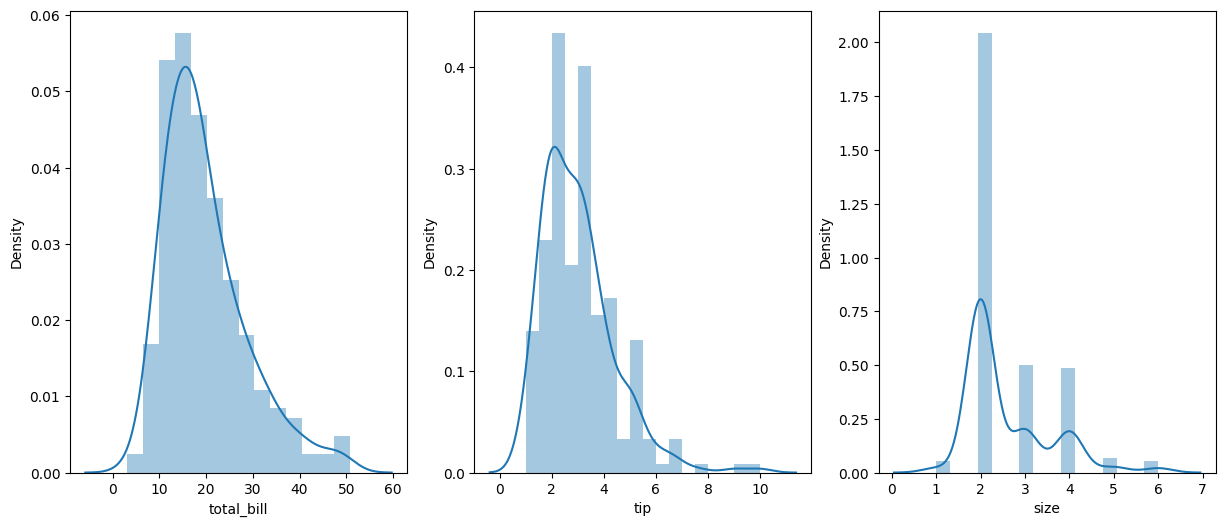
*Gráfica 2.2: Distribuciones variables categóricas*

- Se cree que más hombres que mujeres han dado propina.

- Más personas no fuman, ya sea el que pagó o alguien más del grupo.

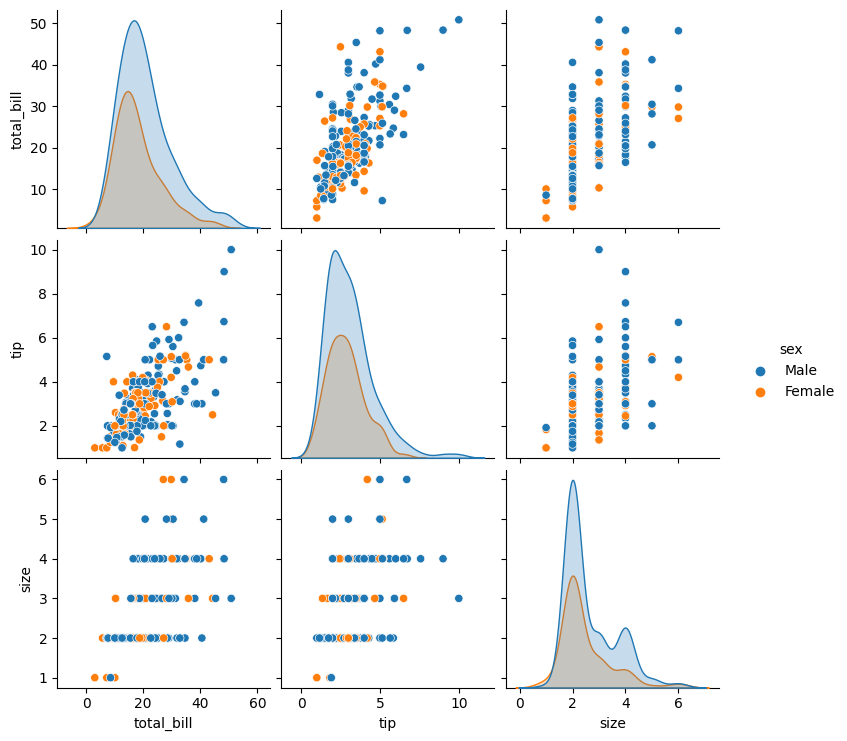
- Solo se trabaja de jueves a domingo. Los viernes hay pocas propinas y la mayor cantidad son los sábados.

- La mayoría de las personas dan propina en la cena.



*Gráfica 2.3: Distribuciones variables numéricas.*

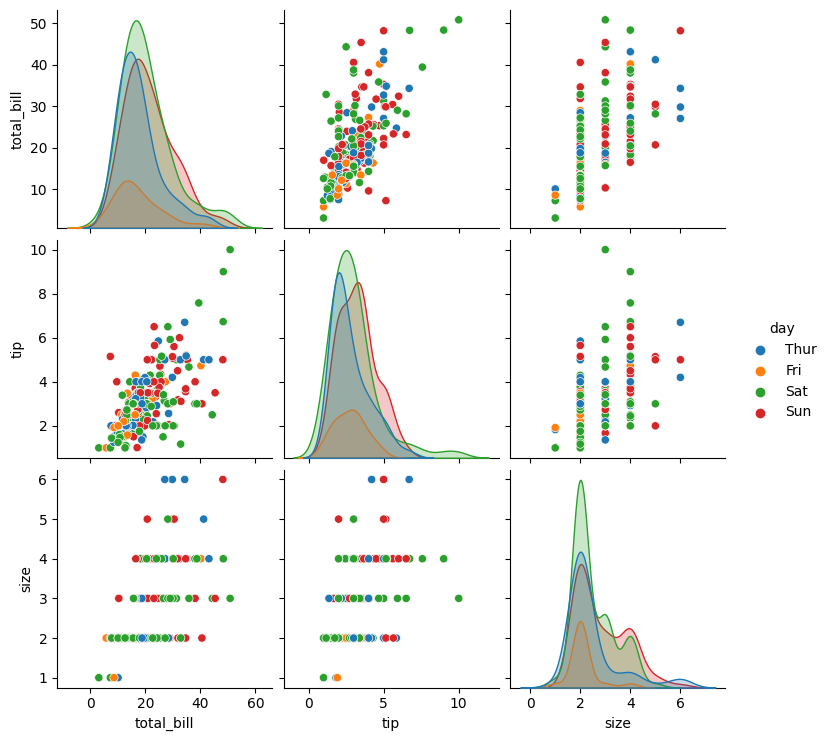
En el caso de la segunda gráfica, se puede apreciar una leve bimodalidad, mientras que las otras tienen caídas más marcadas y tendencias no tan delimitadas. Estas tendencias las separamos en las otras variables categóricas para ver mejor cómo se comporta cada categoría.



*Gráfica 2.4: Scatter plot entre variables y diagramas de densidad de cada variables numéricas. Separado por sexo.*

- La mayoría de los pagos los hacen los hombres pero cuando pagan las mujeres, pagan en la misma proporción (los hombres tienden a pagar las cuentas más altas)

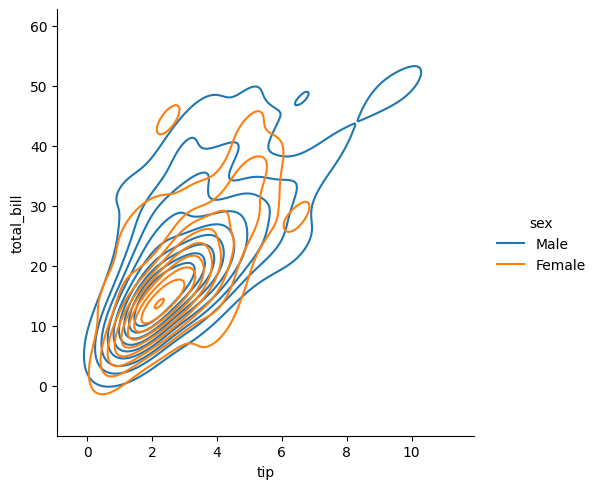
- La tendencia es que vayan 2 personas, pero la mayor cantidad de datos se concentran entre 2 y 4. Es raro que vaya alguien solo.



*Gráfica 2.5: Scatter plot entre variables y diagramas de densidad de cada variables numéricas. Separado por día.*

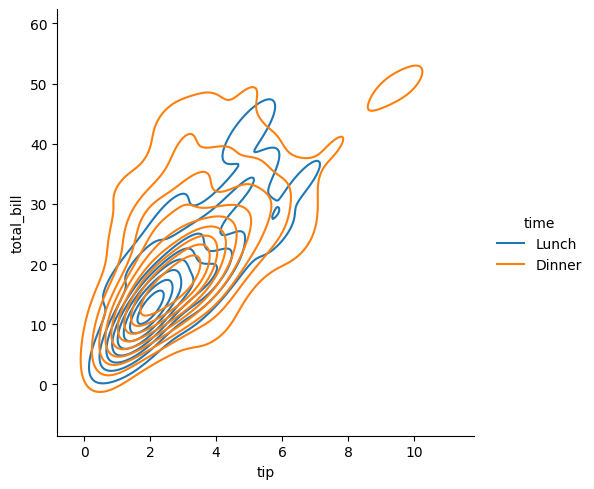
- El sábado es el mejor dia, ya que más personas llegan a consumir y dan mas propinas, en contraste con el viernes que es bastante bajo

- El viernes se puede prescindir de personal, ya que es un dia de bajas visitas (con una subida leve de personal los sábados)



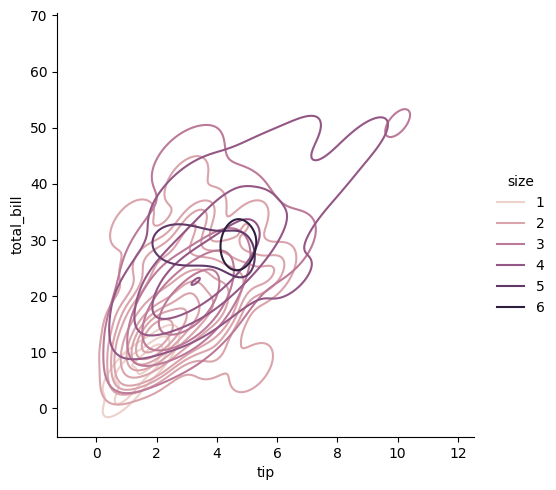
*Gráfica 2.6: Diagrama de calor de total\_bill y tiop. Separado por sexo.*

Las tendencias son idénticas, pero los hombres pagan más cuentas, dan mas propinas y pagan las más caras



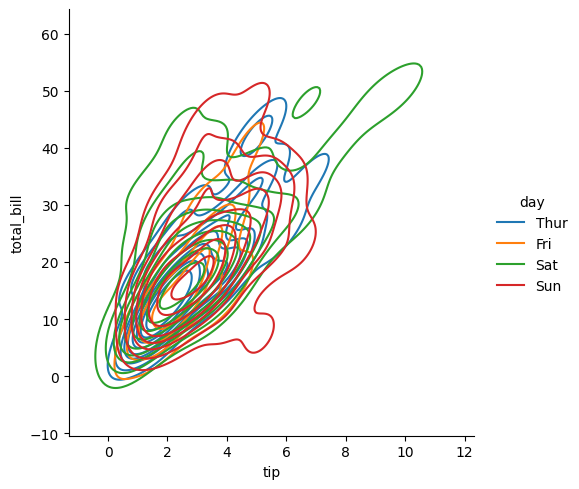
*Gráfica 2.7: Diagrama de calor de total\_bill y tiop. Separado por tiempo.*

La media de pagos en la noche es más alta y más grande que la media de pagos en el almuerzo



*Gráfica 2.8: Diagrama de calor de total\_bill y tiop. Separado por tamaño.*

Las medias de los grupos de personas más grandes tienden a ser las medias de pagos y propinas más grandes, y mientras más grande el grupo, más concentrado está ya que no hay muchas observaciones de grupos grandes.



*Gráfica 2.9: Diagrama de calor de total\_bill y tiop. Separado por día.*

- A pesar de que el viernes las ventas son menores, estas si tienen congruencia con la tendencia general y se amolda a los otros días

- Las medias de cada dia se parecen bastante entre sí

Conclusiones:

* Los restaurantes solo trabajan de jueves a domingo. La mayor actividad se registró el sábado, mientras que el día con menos propinas y compras de comidas fue el viernes. Por esto, podemos usar menos personal los viernes y usar un poco más el sábado (Gráficas 2.2 y 2.5)
* Se puede apreciar una leve bimodalidad en la densidad de las propinas, indicando que hay una doble tendencia para dar propinas. Las otras variables numéricas tienen caídas más marcadas y tendencias no tan delimitadas. (Gráfica 2.3)
* La tendencia es que vayan 2 personas, pero la mayor cantidad de datos se concentran entre 2 personas y 4. Es raro que vaya alguien solo. (Gráfica 2.4)
* Las tendencias de pago entre ambos sexos son idénticas, pero los hombres pagan más veces, dan más propinas y pagan las más caras. (Gráfica 2.6)
* Las medias del número de grupos de personas más grandes tienden a ser las medias de pagos y propinas más grandes, y mientras más grande el grupo, más concentrado está ya que no hay muchas observaciones de grupos grandes. (Gráfica 2.8)