ITESM

MAESTRÍA EN INTELIGENCIA ARTIFICIAL APLICADA

SEGUNDO TRIMESTRE

APUNTES DATA ANALYSIS CON PYTHON

MODULO 3

NOMBRE: VILLALPANDO GUERRERO JIRAM CESAR

Contenido

[**MODULO 3** 2](#_Toc117614790)

[**Graded Review Questions** 2](#_Toc117614791)

# **MODULO 3**

# **Descriptive Statistics**

La estadística descriptiva es, junto con la [inferencia estadística](https://economipedia.com/definiciones/inferencia-estadistica.html) o estadística inferencial, una de las dos grandes ramas de la estadística. Su propio nombre lo indica, trata de describir algo. Pero no describirlo de cualquiera forma, sino de manera cuantitativa. Pensemos en el peso de una caja de verduras, en la altura de una persona o en la cantidad de dinero que gana una empresa.

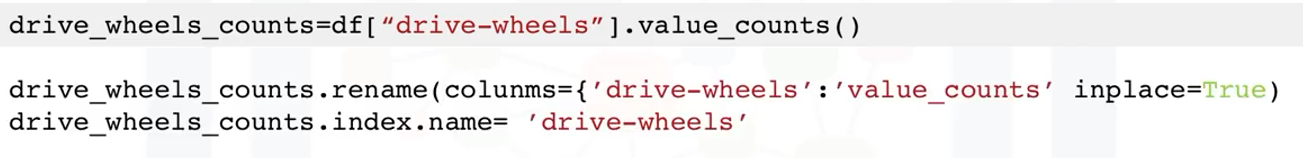
Usando: df.describe()

Nos muestra la media, desviación estándar, valores extremos.

Values\_counts():

Nos ayuda a categorizar la data

Ejemplo:



Box plots: pueden ayudarnos a visualizar de manera gráfica los datos.

Gráfico, Gráfico de cajas y bigotes

Descripción generada automáticamente

En Scatter plos:

1: predecible =X

2: dependiente =y

Ejemplo

Texto

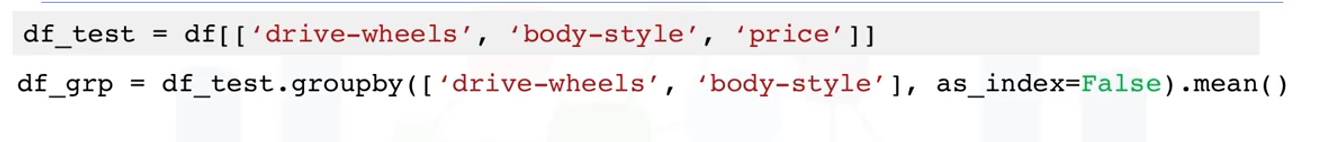
Descripción generada automáticamente

# **GroupBy in Python**

Se puede usar pandas para agrupar en grupos, solos o en múltiples variables.

Dataframe.Groupby()

Ejemplo:



Hay que tener en cuenta el método pivote siguiente:



De esta manera es más fácil visualizarlo

Por otra parte, los mapas de calor, permiten conocer múltiples variables su relación, es muy útil para visualizar intensidades.

Ejemplo;

Imagen que contiene Gráfico de barras

Descripción generada automáticamente

# **Analysis of Variance ANOVA**

Es una fórmula estadística que se utiliza para comparar las varianzas entre las medias (o el promedio) de diferentes grupos. Una variedad de contextos lo utilizan para determinar si existe alguna diferencia entre las medias de los diferentes grupos.

Imagen que contiene Diagrama

Descripción generada automáticamente

Podemos hacer esto a través de:

Diagrama

Descripción generada automáticamente con confianza baja

# **Correlation**

**La correlación es un tipo de asociación entre dos variables numéricas, específicamente evalúa la tendencia (creciente o decreciente) en los datos.**

Dos variables están **asociadas** cuando una variable nos da información acerca de la otra. Por el contrario, cuando no existe asociación, el aumento o disminución de una variable no nos dice nada sobre el comportamiento de la otra variable.

En Python esto se puede usar a través de:



En donde se muestra una correlación positiva ente las dos variables

Gráfico, Gráfico de dispersión

Descripción generada automáticamente

# **Correlation-Statistics**

Algunos métodos para correlación:

* Coeficiente de correlación: entre 1 y -1
* P-value entre .001 y 0.1

Ejemplo:



# **Graded Review Questions**

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente