
**SEMANA 2: ANÁLISIS EXPLORATORIO DE DATOS. ESTADÍSTICA DESCRIPTIVA.
CORRELACIÓN. ANÁLISIS DE VARIANZA**

LABORATORIO CALIFICADO N°2**Indicaciones:**

- El trabajo en su totalidad debe ser hecho en un notebook de 'Jupyter'.
- Cree una cabecera que lleve por título el tema de la semana y debajo de él coloque un 'Markdown' con su nombre y apellido.
- Ejecute su notebook, grábelo, descárguelo como HTML, conviértalo a PDF y suba ambos archivos en el apartado que el docente indique.

Con base a los datos del archivo 'titanic_sin_nan' y tomando como referencia la información en el enlace <https://www.kaggle.com/c/titanic/data>, realice lo siguiente:

1. Cree un data frame, verifique si aún tiene filas con valores nulos y de ser así, elimínelas. Además, seleccione las variables numéricas y realice gráficos de histograma, de densidad, de cajas, de dispersión y un mapa de calor que muestre todas las correlaciones.
2. Haga un análisis para observar el tipo de relación mediante diagramas de dispersión de cada variable numérica con la variable predictora 'fare'. Además, compare y analice cada variable categórica con la variable 'fare' mediante gráficos de cajas y genere mapas de calor por pares de variables categóricas con la misma variable 'fare'.
3. Realice pruebas de hipótesis para analizar si existe causalidad al comparar cada variable numérica con la variable 'fare' mediante el coeficiente de correlación de Pearson tomando un nivel de significancia de 0,05.
4. Realice un ANOVA entre cada variable categórica y la variable 'fare' para analizar si existen diferencias o no entre las medias de los grupos y en caso se rechace la hipótesis nula, haga el análisis por pares de grupos para saber cuáles son significativamente diferentes tomando un nivel de significancia de 0,05.

No olvide convertir su HTML a PDF antes de subirlo. También puede revisar el siguiente enlace para poder hacer esto: <https://mljar.com/blog/jupyter-notebook-pdf/>.