**Universidad Peruana de Ciencias Aplicadas (UPC)**



**Proyecto:**

Anàlisis exploratorio de un conjunto de datos en R/Rstudio

**Curso:**

Administraciuòn de la Informaciòn

**Integrantes:**

* Carhuancho Bazan, Alessandro David u201913933
* Nuñez Lazo, Sergio Antonio u2019102357
* Cruz Mamani, Jack Yefri u201912400

**Ciclo:**

2021-01

**Indice**

1. Caso de anàlisis
2. Conjunto de datos
3. Anàlsis de datos exploratorios
4. Conclusiones preliminares
5. Referencias

**Caso de anàlisis**

El dataset analizado en el presente trabajo fue obtenido desde los repositorios de Kaggle, sin embargo, este dataset ya ha pasado por el proceso de limpieza de datos, el cual fue realizado Thomas Mock y Antoine Bichat para los repositorios de Tidy Tuesday en GitHub, un proyecto semanal enfocado en el análisis de datos usando el lenguaje de programación R. El commit fue realizado el 11 de Febrero del 2020.

El dataset original proviene del artículo Hotel Booking Demand Datasets. Según el artículo, el dataset fue extraído de las bases de datos Property Management System SQL de los dos hoteles examinados. Dichos hoteles se ubican en Portugal, precisamente en el Resort de Algarve, y en la Ciudad de Lisboa.

En el mes de Febrero del año 2019, el artículo fue publicado en el volumen 22 de la revista Data in Brief. Los autores son Nuno Antonio a, b, Ana Almeida a, c, d, y Luis Nunes a, b, d, quienes realizaron estudios en las siguientes instituciones de Portugal:

a. Instituto Universitário de Lisboa, Lisboa

b. Instituto de Telecomunicações, Lisboa

c. CISUC, Coimbra

d. ISTAR-IUL, Lisboa

La utilidad del dato es innegable, sin embargo, la importancia de este dataset en particular se refleja en los siguientes campos:

* Análisis descriptivo para el entendimiento de patrones, tendencias y anomalías en la data.
* Investigación para la predicción de cancelación de una reserva y satisfacción del cliente, dicha información beneficia a los hoteles.
* Benchmarking de algoritmos de clasificación y segmentación, el cual beneficia a los investigadores y educadores de machine learning.
* Entrenamiento de modelos de data mining y estadística para beneficio de los educadores.

**Conjunto de datos (DataSet)**

\* Descripción de la estructura de los datos