

Universidad Internacional de La Rioja

Facultad de Ingeniería y Tecnología

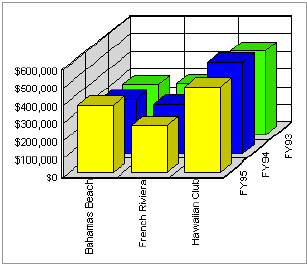
# Máster Universitario en Análisis y Visualización de Datos Masivos / Visual Analitics & Big Data

Visualización de evolución en el tiempo

|  |  |
| --- | --- |
| Actividad de estudio presentado por: | Juan David Escobar Escobar |
| Tipo de trabajo: | Actividad |
| Modalidad: | Individual |
| Profesor/a: | Yusef Hassan Moreno |
| Fecha: | Enero 2022 |

# Actividad

La siguiente gráfica representa la evolución de los ingresos de tres complejos hoteleros (entre 1993 y 1995).



Los datos fuente de esta gráfica son los que aparecen en la siguiente tabla:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Hawaiian Club | French Riviera | Bahamas Beach |
| 1993 | 450000 | 225000 | 245000 |
| 1994 | 475000 | 240000 | 255000 |
| 1995 | 390000 | 205000 | 345000 |

En esta actividad debes:

1. Identificar y describir cuáles los principales errores de diseño que tiene la anterior gráfica:

* El tipo de barras no permiten identificar con exactitud la relación entre las categorías hoteleras y los ingresos obtenidos en cada año (1993, 1994 y 1995)
* La grafica no cuenta con una descripción, leyenda y titulo.
* El eje y indica que el tipo de ingreso es de tipo moneda, pero no indica que unidad de medida o tipo de moneda, por ejemplo (euro, peso o dólar).
* No facilita la interpretación fácil y rápida de la información.

1. Elaborar una visualización que, utilizando los mismos datos como, permita una más intuitiva y eficiente interpretación de estos.

Chart, line chart

Description automatically generated

El grafico de líneas ayuda a la rápida y mejor comprensión, esta grafica se compone de una serie de datos representados por puntos, los cuales están unidos por líneas, se usa por lo general con variables cuantitativas, para estudiar o analizar su comportamiento a lo largo del tiempo. Esta grafica nos permite identificar las tendencias de los datos si van en subida o en bajada y en que año han mejorado o decaído.

El eje X (abscisas) por lo general tiene un valor tipo tiempo, por ejemplo (día, mes, año, semestre, entre otros), el eje Y (ordenadas), en nuestro caso mide el valor que estamos analizando en el tiempo, el cual es los ingresos de los complejos hoteleros de la muestra de datos.

En la parte derecha se complementa la información por medio de dos tipos de herramientas o gráficos tipo tabla, en los cuales se detalla los ingresos totales y los ingresos medios, para cada complejo hotelero a través del tiempo.

El ejercicio de este paso a paso detalla herramientas de visualización, y un conjunto de datos de uso publico adjunto en el siguiente repositorio de [GitHub](https://github.com/juadaves91/bd-bigdata-unir) almacenado en formato CSV.

Para llevar a cabo el proceso de limpieza se opto por utilizar las herramientas que se usan con frecuencia por los analista y científicos de datos, para la recolección, procesamiento y explotación de datos, a continuación, se listas las herramientas y librerías utilizadas:

**Herramientas**: Jupyter Nootebooks Versión 6.4.2 (<https://jupyter.org/>), Python versión 3.0 (<https://www.python.org/>), MongoDB (<https://www.mongodb.com/>), Visual Studio Code (<https://code.visualstudio.com/>), Anaconda Studio (<https://www.anaconda.com/>), Spark versión 3.2.0 (<https://spark.apache.org/>, Atom (<https://atom.io/>).

# Conclusiones

La metodología explicada en este documento se apoya del Análisis Exploratorio de datos, la cual se compone de un conjunto de herramientas, pasos, técnicas estadísticas gráficos, entre otros para llevar a cabo el proceso fundamental para la interpretación básica de los datos y la limpieza de estos.

En este documento se detalla un proceso genérico basado en técnicas existentes y basadas en experiencia en el campo y se evidencian algunos problemas de calidad de datos que se presentan los cuales no son fáciles de tratar, en ocasiones debemos involucrar mas de un rol organizacional y en ocasiones hasta mas de un área organizacional que domina o es responsable una porción del negocio de donde proviene o se consume el dato, no solo esto puede interferir al momento de como manipulamos, limpiamos, estandarizamos o transformamos los datos, también es importante tener presente las reglas de negocio y el propósito o la usabilidad que se le dará a los datos, los escenarios son muy diversos (Dashboards, reportes, Procesos ETL, Bases de datos, informes, entre otros), por lo cual es importante mencionarlo.

# Referencias bibliográficas

datos.gob.es. (22 de 9 de 2021). *datos.gob.es.* Obtenido de Guía Práctica de Introducción al Análisis Exploratorio de Datos: https://datos.gob.es/es/documentacion/guia-practica-de-introduccion-al-analisis-exploratorio-de-datos.

Rocio Chavez Ciencia de Datos. (2020). Histogramas en Python [video]. YouTube.

AIEngineering. (2020). Data Cleaning and Analysis using Apache Spark [video]. YouTube.

Un Analista de Datos Peruano. (2019). Para qué me sirve la Desviación Estandar? [video]. YouTube.

orvizar TV. (2020). Pandas desde cero #3 |Limpiar Datos | Valores nulos | Eliminar datos | tutorial español [video]. YouTube.

Dattos.org. (2019). Gráficos en Python con Matplotlib, Seaborn y Ploty. YouTube.

Tomas\_IA. (2020). Limpieza de datos con Python [video]. YouTube.

Escuela de datos. (28 de 6 de 2017). *Escuela de datos*. Obtenido de Introducción a Pandas y Jupyter Notebook de Python: http://es.schoolofdata.org/2017/06/28/introduccion-a-pandas-y-jupyter-notebook-de-python/

Invarato, R. (20 de 4 de 2021). *jarroba.com*. Obtenido de Instalar Apache Spark en cualquier sistema operativo y aprender a programarlo con Python, Scala o Java: https://jarroba.com/instalar-apache-spark-en-cualquier-sistema-operativo-y-aprender-a-programarlo-con-python-scala-o-java/