



# 1ACC0216-Fundamentos de Data Science

## TB1: Trabajo

---

### I. OBJETIVO

Realizar un análisis exploratorio de un conjunto de datos (EDA) para encontrar patrones de comportamiento, generando visualizaciones, preparando los datos y extrayendo conclusiones iniciales utilizando R/RStudio como herramienta de software.

### II. CONJUNTO DE DATOS

El conjunto de datos motivo de análisis se denomina: Hotel booking demand. Su versión original se obtuvo de Kaggle, sin embargo, para esta evaluación, este conjunto de datos ha sido modificado incorporando ruido en los datos, básicamente: datos faltantes (NA) y datos atípicos (outliers). El conjunto de datos se puede descargar desde [AQUI](#)

En este conjunto de datos se recopilan datos de un hotel urbano y otro de tipo resort. Incluye información de cuándo se realizó la reserva, la duración de la estadía, la cantidad de espacios de estacionamiento disponibles, cantidad de huéspedes adultos, niños y/o bebés, entre otros datos.

El conjunto de datos original proviene del documento: [Hotel booking demand datasets](#)

### III. DOCUMENTO ENTREGABLE

El grupo de estudiantes entregará un único documento de acuerdo a la nomenclatura de archivos, desarrollando los siguientes temas en este orden propuesto:

#### 1. CASO DE ANALISIS

- **Origen de los Datos:**

Breve explicación sobre la procedencia de los datos: fuente, autor(es), fecha, país o región. Incluir una mención de la credibilidad de los datos y cómo fueron recolectados (encuestas, registros administrativos, etc.).

- **Casos de Uso Applicable:**

- ¿Quién podría estar interesado en este análisis? (p. ej., hoteles, agencias de viajes, departamentos de marketing, etc.)
- ¿Qué problemas o necesidades responde este análisis? (p. ej., optimización de ocupación, predicción de demanda, toma de decisiones en marketing).



# 1ACC0216-Fundamentos de Data Science

## TB1: Trabajo

---

### 2. CONJUNTO DE DATOS (DATA SET)

- **Descripción del Data Set:**

Proporcionar una tabla que describa las variables contenidas en el conjunto de datos, especificando:

- Nombre de la variable.
- Tipo de dato (numérico, categórico, etc.).
- Descripción breve del significado de cada variable.

Ejemplo:

Variable	Tipo	Descripción
hotel	Categórico	Tipo de hotel (Resort Hotel, City Hotel)
arrival_date	Fecha	Fecha de llegada al hotel
children	Numérico	Número de niños incluidos en la reserva

### 3. ANÁLISIS EXPLORATORIO DE DATOS (EDA)

Descripción de instrucciones ejecutadas en R/RStudio y resultados obtenidos para:

- ❖ **CARGAR DATOS**

Código y procedimiento utilizado para cargar los datos en R/RStudio, asegurándose de incluir los parámetros necesarios (header = TRUE, stringsAsFactors = FALSE).

- ❖ **INSPECCIONAR DATOS**

Mostrar las instrucciones y salidas que permitan comprender la estructura del conjunto de datos, identificando los distintos tipos y formatos de variables presentes (como numéricas, categóricas, fechas, entre otros) y duplicados.

Realizar los ajustes y transformaciones a las variables según sea necesario para asegurarnos de que se puedan utilizar eficazmente en el análisis.



# 1ACC0216-Fundamentos de Data Science

## TB1: Trabajo

---

### ❖ PRE-PROCESAR DATOS

- **Resumir Estadísticas Básicas:**  
Este resumen inicial nos dará una idea general del comportamiento de las variables.
- **Identificación de Datos Faltantes:**  
Proceso para identificar valores NA en el conjunto de datos.
- **Tratamiento de Datos Faltantes:**  
Técnica utilizada para eliminar o completar datos faltantes (p. ej., imputación, eliminación).
- **Detectar Outliers:**  
Procedimiento para detectar valores atípicos, como diagramas de caja (boxplots).
- **Tratamiento de Outliers:**  
Técnica(s) para transformar o manejar los valores atípicos (p. ej., winsorización).  
Guardar el archivo para realizar el siguiente paso

### ❖ VISUALIZACION DE DATOS

El análisis debe dar respuesta a las siguientes preguntas a través de visualizaciones basadas en los datos pre-procesados:

- ¿Cuántas reservas se realizan por tipo de hotel? ¿Qué tipo de hotel prefiere la gente?
  - ¿Está aumentando la demanda con el tiempo?
  - ¿Cuáles son las temporadas de reservas (alta, media, baja)?
  - ¿Cuál es la duración promedio de las estancias por tipo de hotel?
  - ¿Cuántas reservas incluyen niños y/o bebés?
  - ¿Es importante contar con espacios de estacionamiento?
  - ¿En qué meses del año se producen más cancelaciones de reservas?
  - Plantear una pregunta del equipo
- 
- Justificación de las visualizaciones elegidas (gráficos de líneas, barras, diagramas de dispersión, etc.).
  - Hallazgos obtenidos mediante las visualizaciones.



# 1ACC0216-Fundamentos de Data Science

## TB1: Trabajo

---

### IV. CONCLUSIONES

- **Conclusiones Basadas en el Análisis:**

Resumen de las respuestas obtenidas para cada una de las preguntas planteadas en el caso de análisis.

- ¿Qué patrones o tendencias se observaron?
- ¿Qué recomendaciones se pueden extraer a partir de los hallazgos?

### . Recomendaciones:

- **Claridad en las Visualizaciones:**

Las visualizaciones deben ser fáciles de interpretar y deben estar bien etiquetadas (títulos, leyendas, ejes). Si algunos ameritan mostrar tablas o gráficos diferenciado por tipo de hotel (urbano/resort)

- **Interpretación Basada en Datos:**

Asegurar que las conclusiones estén respaldadas por el análisis de los datos y las visualizaciones presentadas.

### V. ARCHIVAR Y PUBLICAR

- Se deberá contemplar un repositorio en Github.com llamado: 1ACC0216- TB1-2025-1 conteniendo dos carpetas:
  - **data:** deberá contener el dataset original y el dataset final resultante (limpio o preparado para análisis).
  - **code:** deberá contener los scripts en R utilizados para el proceso de carga, inspección, preprocesado y visualización del dataset.
- El archivo Readme, dentro de GitHub, deberá contemplar:
  - Objetivo del trabajo
  - Nombre de los alumnos participantes
  - Breve descripción del dataset (se puede adjuntar el archivo PDF)
  - Conclusiones
  - Licencia

Guiarse de estos ejemplos de publicaciones de trabajos en GitHub:



# 1ACC0216-Fundamentos de Data Science

## TB1: Trabajo

---

<https://github.com/fernandoabcampos/titanic-data-cleaning-and-validation>

- (Opcional) El mismo contenido publicado en GitHub reproducirlo en la sección Wiki perteneciente al grupo en el Aula Virtual.
- En el documento entregable, se deberá incluir el enlace a la cuenta de GitHub.com desde donde se accede a la publicación de la evaluación.

### **Nomenclatura de Archivos:**

upc-nrogrupo-tb1 (.docx o .pdf)

upc-nrogrupo-tb1.R

upc-nrogrupo-tb1(pptx u otros tipos de presentación)

Sólo se debe entregar tres archivos por grupo.

### **Evaluación:**

Se evaluará de acuerdo con la siguiente rúbrica:

### **Consideraciones adicionales:**

- Se evaluará el orden dentro de la organización del documento, así como la correcta redacción y gramática.
- Durante la semana 7 se realizará la exposición grupal.
- Cada grupo contará con 10-15 minutos para que cada integrante detalle cómo se obtuvieron las visualizaciones y conclusiones (preparar una presentación de no más de diez diapositivas).
- El orden de la exposición será mediante sorteo.
- El alumno que no se presente, perderá la calificación oral.
- La calificación podrá ser diferenciada por alumno si su participación /respuestas durante la exposición es insuficiente.