

Identificación de tendencias en la duración de los vuelos en los viajes aéreos

Exploración de datos de vuelos desde Nueva York mediante técnicas avanzadas de unión de datos en R.

Descripción del Proyecto

Vuele al mundo del análisis de datos de aerolíneas, centrándose en los vuelos desde Nueva York en la segunda mitad de 2022. Mejore sus habilidades de manipulación de datos explorando las duraciones, frecuencias y destinos de los vuelos utilizando tidyverse y obtenga información valiosa sobre las operaciones de las aerolíneas.

1. ¿Qué par de aerolíneas y aeropuertos recibe la mayor cantidad de vuelos desde Nueva York y cuál es la duración promedio de ese vuelo?
2. Guarde su respuesta como un conjunto de datos llamado frecuente con una fila y un mínimo de dos columnas: **airline_name**, **aeroportuary_name**.
3. Encuentre el aeropuerto que tiene la duración promedio de vuelo más larga (en horas) desde Nueva York. ¿Cómo se llama este aeropuerto?
4. Guarde su respuesta como un conjunto de datos llamado más largo con una fila y un mínimo de dos columnas: **airline_name**, **aeroportuary_name**.
5. ¿Cuál es el aeropuerto con el destino menos frecuentado por los vuelos que salen de JFK?
6. Guarde su respuesta como una cadena de caracteres llamada **least**.

Los datos

Un actor destacado de la industria de la aviación con una presencia significativa en la ciudad de Nueva York ha lanzado un proyecto de análisis de datos en profundidad centrado en la identificación de tendencias en la duración de los vuelos en los viajes aéreos. Esta iniciativa tiene como objetivo profundizar en una gran cantidad de datos relacionados con los horarios de los vuelos y los patrones operativos, con el objetivo de optimizar los tiempos de vuelo y mejorar la experiencia de viaje general para los pasajeros. Como analista de datos principal, tiene acceso a conjuntos de datos enriquecidos, provenientes de la colección **'nycflights2022'** producida por el equipo de ModernDive. Estos conjuntos de datos incluyen registros de vuelos que salen de los principales aeropuertos de la ciudad de Nueva York, incluidos JFK (Aeropuerto Internacional John F. Kennedy), LGA (Aeropuerto LaGuardia) y EWR (Aeropuerto Internacional Newark Liberty), durante la segunda mitad de 2022. Ofrecen una visión integral de las operaciones de vuelo, que abarca varios aspectos como los horarios de salida y llegada, las rutas de vuelo y los detalles de las aerolíneas:

Conjunto de Datos: **flights2022-h2.csv**

Columna	Descripción
carrier	Código de la compañía aérea
origin	Aeropuerto de origen (código IATA)
dest`	Aeropuerto de destino (código IATA)
air_time	Duración del vuelo en el aire, en minutos

Conjunto de Datos: **airlines.csv**

Columna	Descripción
carrier	Código de la aerolínea
name	Nombre completo de la aerolínea

Conjunto de Datos: **airports.csv**

Columna	Descripción
faa	Código FAA del aeropuerto
name	Nombre completo del aeropuerto

Agenda de Actividades del Proyecto

1. Unión y resumen de datos complejos (25 puntos)
2. Búsqueda del destino de vuelo más frecuente (25 puntos)
3. Determinación de la duración más larga del vuelo (25 puntos)
4. Descubrimiento del destino menos común (25 puntos)

1. UNION Y RESUMEN DE DATOS COMPLEJOS

Debes unir (join) los conjuntos de datos de vuelos, aerolíneas y aeropuertos, calcula la duración del vuelo y analiza la duración promedio del vuelo y la cantidad de vuelos para cada combinación de aerolínea y aeropuerto.

Instrucciones Paso 1: Unión y resumen de datos complejos (25 puntos)

Unir conjuntos de datos

Use **left_join()** para fusionar vuelos con aerolíneas por "transportista" y aeropuertos por "destino" ("faa" en aeropuertos).

Cambie el nombre de la columna de nombre a **airline_name** y **aeroportuaria_name** después de la fusión.

Cálculo de la duración del vuelo

Utilice **mutate()** en **air_time** para agregar una columna para la duración del vuelo en horas.

Divida **air_time** por 60 para obtener el tiempo en horas.

Análisis de la duración y el conteo promedio de los vuelos

Agrupe por **nombre_de_la_aerolínea** y **nombre_del_aeropuerto**, luego, **summarize()** la duración y el conteo promedio de los vuelos.

Ignore los valores faltantes estableciendo **na.rm = TRUE**.

Asegúrese de eliminar la estructura de agrupamiento al final.

2. ENCONTRAR EL DESTINO DE VUELO MAS FRECUENTE.

Analiza qué aerolínea y ciudad tienen más vuelos desde Nueva York ordenando los datos.

Instrucciones Paso 2: Encuentra el destino de vuelo más frecuente (25 puntos)

Cómo organizar los datos

Puedes usar una combinación de **organize()** y **desc()** para ordenar los datos de conteo de vuelos que tienes.

Identificación del valor más frecuente

Si se organiza u ordena correctamente, la fila superior debe contener el valor más frecuente.

Encadenar con **head()** para extraer la primera fila del marco de datos y almacenarla como **frequent**.

3. COMO DETERMINAR LA DURACION DE VUELO MAS LARGA.

Identifique qué aerolínea y aeropuerto tienen la duración de vuelo promedio más larga desde Nueva York organizando los datos de otra manera.

Instrucciones Paso 3: Cómo determinar la duración de vuelo más larga (25 puntos)

Ordenar datos

Puede utilizar una combinación de **organize()** y **desc()** para ordenar los datos de vuelo promedio

Encontrar la duración de vuelo promedio más larga

Si se organiza u ordena correctamente, la fila superior debe contener la duración de vuelo promedio más larga.

Encadenar con **head()** para extraer la primera fila del conjunto de datos y almacenarla como la más larga.

4. DESCUBRA EL DESTINO MENOS COMUN.

Descubra el aeropuerto de destino menos común para los vuelos que salen del aeropuerto JFK analizando un subconjunto agrupado de los datos.

Instrucciones Paso 4: Descubra el destino menos común (25 puntos)

Filtrar los datos

Puede crear un subconjunto de los datos con la función **filter()**.

Deberá conservar los datos cuyo origen sea "JFK".

Agrupar los datos

Utilice **group_by()** para agrupar los datos por nombre de aeropuerto.

Analiza el subconjunto agrupado

Encuentra el conteo de cada grupo con **summary()**.

Ordena por el conteo con **arrangement()** para encontrar el destino menos común.

Guarda tu respuesta como una cadena de caracteres llamada **least**.