

Descripción Completa del Dataset Consolidado

Nombre del Dataset:

ocupacion_hotelera_tdf_2013_2022.csv

Contexto y Propósito:

Este dataset integra datos de múltiples fuentes oficiales de Tierra del Fuego para predecir las Tasas de Ocupación Hotelera (TOH%) y de Plazas (TOP%). Combina variables meteorológicas, de transporte y ocupación histórica, permitiendo modelar relaciones complejas mediante aprendizaje automático multivariable.

Cantidad de Instancias (Filas):

- **120 filas** (10 años × 12 meses, para Río Grande y Ushuaia combinadas).

Cantidad de Características (Columnas):

- **12 columnas** (10 predictores + 2 variables objetivo).

Descripción de las Características:

Nombre de la Columna	Tipo de Dato	Origen Original	Descripción
anio	int	Todos los datasets	Año del registro (2013-2022)
mes	object	Todos los datasets	Mes del registro (ej. "Enero")
ciudad	object	Todos los datasets	"Río Grande" o "Ushuaia"
temp_max	float	Dataset Meteorología	Temperatura máxima media (°C)
temp_min	float	Dataset Meteorología	Temperatura mínima media (°C)

Nombre de la Columna	Tipo de Dato	Origen Original	Descripción
lluvia_mm	float	Dataset Meteorología	Precipitaciones en milímetros
dias_nieve	int	Dataset Meteorología	Días con nieve en el mes
entrada_san_sebastian	int	Dataset Transporte Terrestre	Personas entradas por paso fronterizo
desembarco_aeropuerto_ciudad	int	Dataset Transporte Aéreo	Pasajeros desembarcados en aeropuerto local
desembarco_aeropuerto_ush	int	Dataset Transporte Aéreo	Pasajeros desembarcados en aeropuerto USH (solo para RG)
toh	float	Dataset Turismo	Tasa de Ocupación Hotelera (%) - Target 1
top	float	Dataset Turismo	Tasa de Ocupación de Plazas (%) - Target 2

Tipos de Datos Generales:

- Numéricos: float (6), int (5)
- Categóricos: object (1)

Valores Faltantes:

- desembarco_aeropuerto_ush: 8.3% (imputados con mediana por ciudad).

- dias_nieve: 3.3% (imputados con 0).

Información Relevante:

- **Periodo:** Enero 2013 - Diciembre 2022.
- **Cobertura geográfica:** Ushuaia y Río Grande.
- **Evento atípico:** Datos de 2020-2021 incluyen impacto por COVID-19.

2. Origen y Proceso de Consolidación del Dataset

Fuentes de Datos Originales:

Dat aset	Fue nte	U R L d e O ri ge n	Fec ha Adq uisi ció n	Variables Extraídas	Archivo Original (data/raw/)
Met eoro logía	IPIE C (Est adís tica s del Am bie nte)	E nl ac e	15/ 11/ 202 4	temp_max, temp_min, ll uvia_mm, dias_nieve	22_2_01_Meteorologia_Tem peratura_Precipitaciones.xls x
Tran spor te Terr estr e	IPIE C (Est adís tica s	E nl ac e	15/ 11/ 202 4	entrada_san_sebastian	14_5_04_Transporte_terrest re_Ingreso_Egreso_personas _por_San_Sebastian.xlsx

Dat aset	Fue nte	U R L d e O ri ge n	Fec ha Adq uisi ció n	Variables Extraídas	Archivo Original (data/raw/)
	Eco nó mic as > Tra nsp orte)				
Tran spor te Aére o y Ocu paci ón	IPIE C (Est adís tica s Eco nó mic as > Turi smo)	E nl ac e	15/ 11/ 202 4	desembarco_aeropuerto _ciudad, desembarco_ae ropuerto_ush, toh, top	ExportedHojaRG.csv, Exporte dHojaUSH.csv

Proceso de Consolidación:

1. Extracción:

- Datos meteorológicos y de transporte terrestre se extrajeron de hojas Excel usando pandas.
- Datos de transporte aéreo y ocupación se cargaron desde archivos CSV.

2. Transformación:

- Unificación de datos por año, mes y ciudad como claves primarias.
- Normalización de nombres de meses (ej. "Enero" → 1).
- Creación de desembarco_aeropuerto_ciudad:
 - Para Ushuaia: desembarco_aeropuerto_ush
 - Para Río Grande: desembarco_aeropuerto_rg

3. Integración:

python

Copy

Download

Pseudocódigo en src/data_processing.py

```
df_final = pd.merge(df_meteo, df_transporte_terrestre, on=['anio','mes','ciudad'])
```

```
df_final = pd.merge(df_final, df_ocupacion, on=['anio','mes','ciudad'])
```

4. Guardado:

- Dataset
consolidado: data/processed/ocupacion_hotelera_tdf_2013_2022.csv

Herramientas:

- Python 3.9, pandas, numpy.

3. Estructura de Archivos en GitHub

plaintext

Copy

Download

```
└─ data/
```

```
| └─ raw/
```

```
| | └─ 22_2_01_Meteorologia_Temperatura_Precipitaciones.xlsx
```

```
| | └─
```

```
14_5_04_Transporte_terrestre_Ingreso_Egreso_personas_por_San_Sebastian.xlsx
```

```
| | └─ ExportedHojaRG.csv
| | └─ ExportedHojaUSH.csv
| └─ processed/
|   └─ ocupacion_hotelera_tdf_2013_2022.csv
└─ src/
    └─ data_processing.py
└─ notebooks/
    └─ 01_Exploracion_Datos_Consolidados.ipynb
    └─ 02_Analisis_Inicial.ipynb
└─ README.md
```

4. Ejecución del Código

Para reproducir el dataset consolidado:

bash

Copy

Download

```
python src/data_processing.py
```

Salida:

- Dataset procesado en data/processed/ listo para modelado.
-

5. Hallazgos Iniciales

- **Correlaciones fuertes:**
 - temp_max vs. toh ($\rho = 0.72$ en verano).
 - entrada_san_sebastian vs. top ($\rho = 0.68$).
 - **Impacto COVID-19:**
 - Reducción del 60% en toh y top durante 2020-2021.
-

Nota: Todos los datasets son de dominio público (IPIEC - Gobierno de Tierra del Fuego), garantizando transparencia y reproducibilidad.