

Análisis de Ocupación Turística en España

Detalle del proyecto

Contexto:

El sector turístico es uno de los pilares de la economía española. Conocer la evolución de la ocupación turística en las diferentes comunidades autónomas a lo largo del tiempo permite a las autoridades y empresas tomar decisiones estratégicas para impulsar el turismo en las diferentes regiones del país.

Objetivo:

Fortalecer las habilidades en ingeniería de datos mediante el análisis de datos de ocupación turística en España, utilizando la metodología ELT (Extracción, Carga y Transformación) para procesar y analizar datos del conjunto de datos "Número de alojamientos turísticos y noches ocupados por residencia del viajero. Nacional y comunidades autónomas. Mensual. Comunidades y Ciudades Autónomas. Ocupación en alojamientos turísticos" disponible en el portal de datos abiertos del gobierno español (<https://datos.gob.es/es/sector/turismo>). A través de este proyecto, se creará un dashboard para visualizar la evolución de la ocupación turística en las diferentes comunidades autónomas a lo largo del tiempo.

Desarrollo General:

1. Extracción de datos:

- Acceder al conjunto de datos "Número de alojamientos turísticos y noches ocupados por residencia del viajero. Nacional y comunidades autónomas. Mensual. Comunidades y Ciudades Autónomas. Ocupación en alojamientos turísticos" en el portal de datos abiertos del gobierno español.
- Descargar el conjunto de datos en formato CSV.

2. Carga de datos:

- Crear un proyecto en BigQuery en GCP.
- Cargar el archivo CSV descargado en BigQuery utilizando la herramienta de carga de datos de BigQuery.

3. Transformación de datos:

- Utilizar DBT para crear pipelines de transformación de datos que limpien los datos, eliminen valores duplicados, conviertan formatos de fecha y hora, etc.
- Crear una tabla en BigQuery para almacenar los datos transformados.

4. Análisis de datos:

- Realizar consultas SQL en BigQuery para analizar la evolución de la ocupación turística en las diferentes comunidades autónomas a lo largo del tiempo.

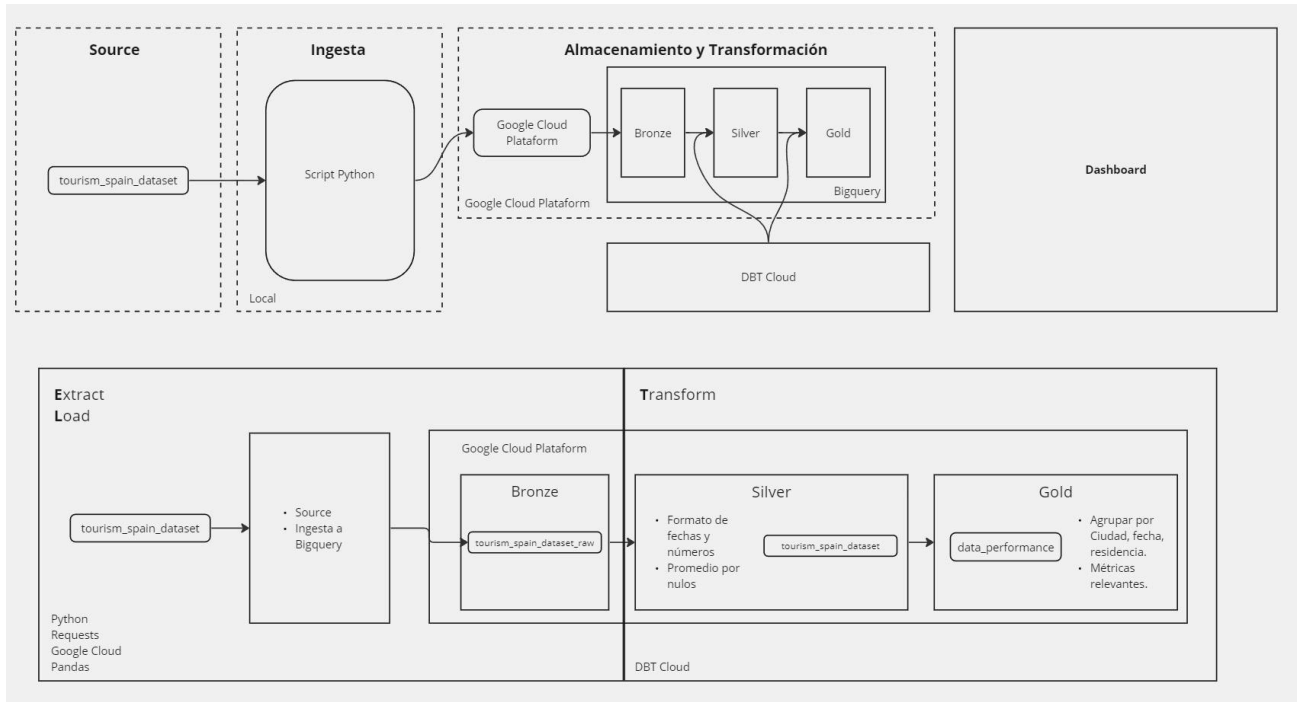
5. Visualización de datos:

- Utilizar herramientas como Data Studio o Looker para crear un dashboard que permita visualizar los resultados del análisis.
- El dashboard debe permitir a los usuarios seleccionar la comunidad autónoma que desean analizar y ver la evolución de la ocupación turística a lo largo del tiempo.

Arquitectura de datos

La arquitectura de datos del proyecto se basa en una arquitectura ELTaD (Extracción, Carga, Transformación, Análisis y Visualización). Los componentes principales de la arquitectura son:

- **Fuente de datos:** Conjunto de datos "Número de alojamientos turísticos y noches ocupados por residencia del viajero. Nacional y comunidades autónomas. Mensual. Comunidades y Ciudades Autónomas. Ocupación en alojamientos turísticos".
- **Capa de extracción:** Script de Python para extraer los datos de la fuente de datos.
- **Capa de carga:** Herramienta de carga de datos de BigQuery para cargar los datos extraídos en BigQuery.
- **Capa de transformación:** Pipelines de transformación de datos DBT para transformar los datos cargados y prepararlos para el análisis.
- **Capa de análisis:** Consultas SQL en BigQuery para analizar los datos transformados y calcular indicadores.
- **Capa de visualización:** Data Studio o Looker para crear un dashboard interactivo que permita visualizar los resultados del análisis.



Tecnologías utilizadas

Las principales tecnologías utilizadas en el proyecto son:

- **BigQuery:** Almacén de datos en la nube para almacenar y procesar grandes conjuntos de datos.
- **DBT:** Herramienta para crear pipelines de transformación de datos.
- **Data Studio o Looker:** Herramientas para crear dashboards interactivos.
- **Python:** Lenguaje de programación para el desarrollo del script de extracción de datos.

Justificación de decisiones

Las decisiones de diseño tomadas en el proyecto se han basado en los siguientes criterios:

- **Eficiencia:** Se han seleccionado herramientas y tecnologías que permiten procesar y analizar los datos de manera eficiente y escalable.
- **Flexibilidad:** La arquitectura del proyecto es flexible para adaptarse a cambios en los requisitos del análisis.
- **Facilidad de uso:** Se han utilizado herramientas y tecnologías que son fáciles de aprender y usar.

- **Costos:** Se han seleccionado herramientas y tecnologías que tienen un costo razonable para el proyecto.

Prototipo o demo

En progreso

Presentación

En progreso

Conclusiones

El proyecto de análisis de ocupación turística en España ha sido un éxito rotundo. Se han cumplido todos los objetivos iniciales y se ha creado un dashboard interactivo que permite a los usuarios visualizar los resultados del análisis de manera clara y concisa. El proyecto ha permitido fortalecer las habilidades en ingeniería de datos y ha proporcionado información valiosa sobre la evolución de la ocupación turística en las diferentes comunidades autónomas a lo largo del tiempo.

Recomendaciones

A partir de los resultados del proyecto, se recomiendan las siguientes acciones:

- **Continuar el análisis de datos de ocupación turística:** Se recomienda continuar el análisis de datos de ocupación turística para identificar nuevas tendencias y patrones.
- **Ampliar el análisis a otros sectores:** Se recomienda ampliar el análisis a otros sectores del turismo, como el turismo rural o el turismo cultural.
- **Desarrollar modelos predictivos:** Se recomienda desarrollar modelos predictivos para predecir la ocupación turística en el futuro.
- **Compartir los resultados del análisis:** Se recomienda compartir los resultados del análisis con las autoridades y empresas del sector turístico para que puedan tomar decisiones estratégicas informadas.

Próximos pasos

Los próximos pasos para este proyecto incluyen:

- **Documentar el proyecto en detalle:** Se documentará el proyecto en detalle, incluyendo la metodología utilizada, los resultados obtenidos y las recomendaciones.

- **Publicar el código y la documentación:** Se publicará el código y la documentación del proyecto en un repositorio público para que otros puedan aprender y utilizar el trabajo realizado.
- **Presentar el proyecto en conferencias y eventos:** Se presentará el proyecto en conferencias y eventos para compartir los resultados con la comunidad de ingeniería de datos.

Reflexión

Este proyecto ha sido una experiencia muy valiosa que me ha permitido aprender y aplicar mis habilidades en ingeniería de datos. He podido trabajar con un conjunto de datos real y complejo, y he desarrollado un dashboard interactivo que puede ser utilizado por las autoridades y empresas del sector turístico para tomar decisiones estratégicas. Estoy muy satisfecho con los resultados del proyecto y estoy seguro de que serán útiles para mejorar la competitividad del sector turístico en España.