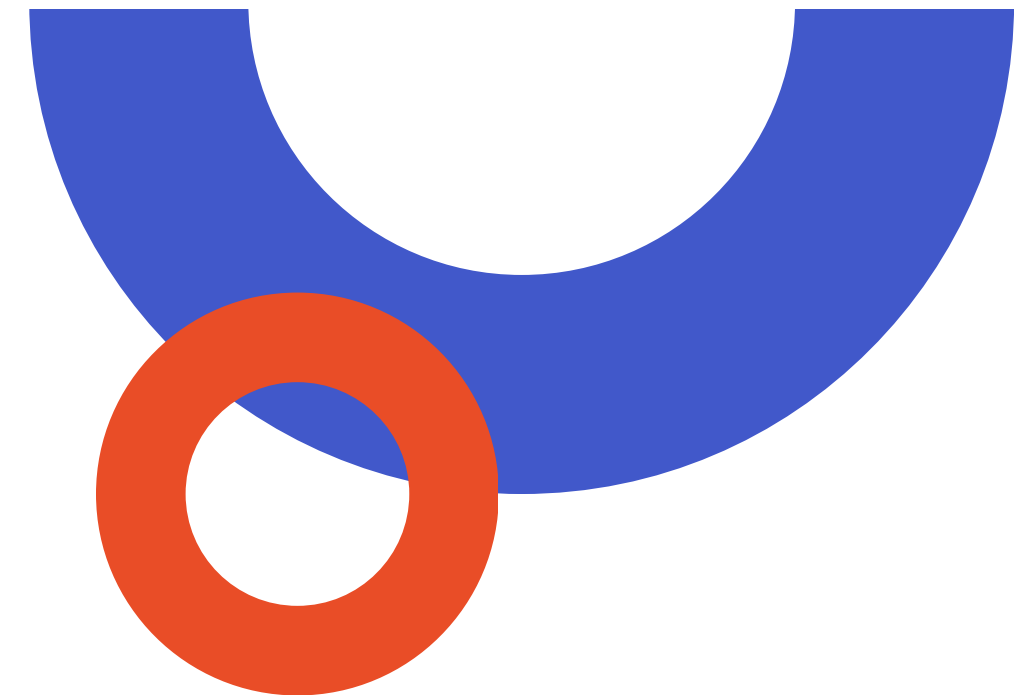


# PROYECTO DE ANÁLISIS DE DATOS

## INTEGRANTES

- Jordy Navarro
- José Vargas
- Carlos Quintana
- Joshua Morocho
- Cristian Tambaco



# INDICE DE CONTENIDOS

- 01. Objetivos
- 02. Temas elegidos
- 03. Definición de casos de estudio
- 04. Herramientas utilizadas
- 05. Arquitectura de solución
- 06. Visualización de Datos
- 07. Conclusiones y Recomendaciones



# Objetivos

Aplicar los conocimientos adquiridos sobre extracción, limpieza, transformación, visualización y análisis de datos mediante Power BI o Python.



# TEMAS ELEGIDOS

## EVENTOS DEPORTIVOS A NIVEL MUNDIAL

Asignado a: Jose Vargas

## RESTAURANTES Y SITIOS DE ESPARCIMIENTO

Asignado a: Carlos Quintana

## ACTIVIDADES Y HOBBIES

Asignado a: Cristian Tambaco

## EVENTOS O NOTICIAS MUNDIALES

Asignado a: Joshua Morocho

## ACCIDENTES DE TRANSITO DESDE EL 2002 HASTA 2024 A NIVEL MUNDIAL

Asignado a: Jordy Navarro



# Definición de casos de estudio



- Análisis de tendencias en el rendimiento deportivo global
- Objetivo: Analizar el rendimiento de diferentes países en eventos deportivos internacionales.
- Fuente de Datos: Eventos deportivos a nivel mundial.
- Impacto de los accidentes de tránsito en la movilidad urbana
- Objetivo: Estudiar cómo los accidentes de tránsito afectan la congestión y la movilidad en ciudades clave a nivel mundial.
- Fuente de Datos: Accidentes de tránsito desde 2002 hasta 2024.
- Relación entre los hobbies y el bienestar
- Objetivo: Estudiar cómo los diferentes hobbies están relacionados con indicadores de bienestar.
- Fuente de Datos: Actividades y hobbies.

# Definición de casos de estudio



- Desempeño de los restaurantes durante festividades globales
  - Objetivo: Examinar cómo las festividades globales (como Navidad, Año Nuevo, etc.) impactan la demanda de restaurantes y sitios de esparcimiento.
  - Fuente de Datos: Restaurantes y sitios de esparcimiento, Eventos mundiales.
- 
- Efecto de los eventos mundiales en la disponibilidad de recursos en restaurantes
  - Objetivo: Estudiar cómo los eventos mundiales afectan la disponibilidad de productos y la demanda en restaurantes.
  - Fuente de Datos: Eventos mundiales, Restaurantes y sitios de esparcimiento.

# Herramientas utilizadas.



## **MYSQL**

Es una base de datos relacional, MySQL organiza los datos en tablas con filas y columnas, siguiendo un esquema fijo que se define al principio. Cada tabla puede tener relaciones con otras a través de claves primarias y claves foráneas.



## **MONGODB COMPASS**

MongoDB es una base de datos NoSQL, no utiliza el modelo de datos relacional tradicional (como las bases de datos SQL). Está diseñada para almacenar grandes volúmenes de datos no estructurados o semi-estructurados.

# Herramientas utilizadas.



**CouchDB**

## **COUCHDB**

CouchDB es un sistema de gestión de bases de datos NoSQL, de código abierto. Almacena la información en documentos en formato JSON. Es accesible mediante una API RESTful. Se interactúa con la base de datos usando solicitudes HTTP comunes (como GET, POST, PUT y DELETE)



Microsoft®  
**SQL Server**®

## **SQL SERVER**

Utiliza el modelo relacional, donde los datos se almacenan en tablas, cada una con filas y columnas. Estas tablas pueden estar relacionadas entre sí mediante claves primarias y claves foráneas.



# Arquitectura de solución.

## A. Bases de Datos Relacionales:

1. SQL Server: Como base de datos final donde se almacenará la información consolidada, debido a que es el repositorio final que se requiere.

- En esta base de datos se almacenarán los registros principales y los datos procesados que se integren desde las fuentes.
- También es donde se realizarán los análisis iniciales y consultas complejas de datos.

## B. Bases de Datos NoSQL:

1. MongoDB: Usada para almacenar datos semi-estructurados como JSON. MongoDB es excelente para manejar grandes volúmenes de datos no estructurados.

2. CouchDB: Utilizada para almacenar y gestionar documentos en formato JSON. CouchDB es especialmente buena para aplicaciones que requieren alta disponibilidad y escalabilidad.

# Arquitectura de solución.

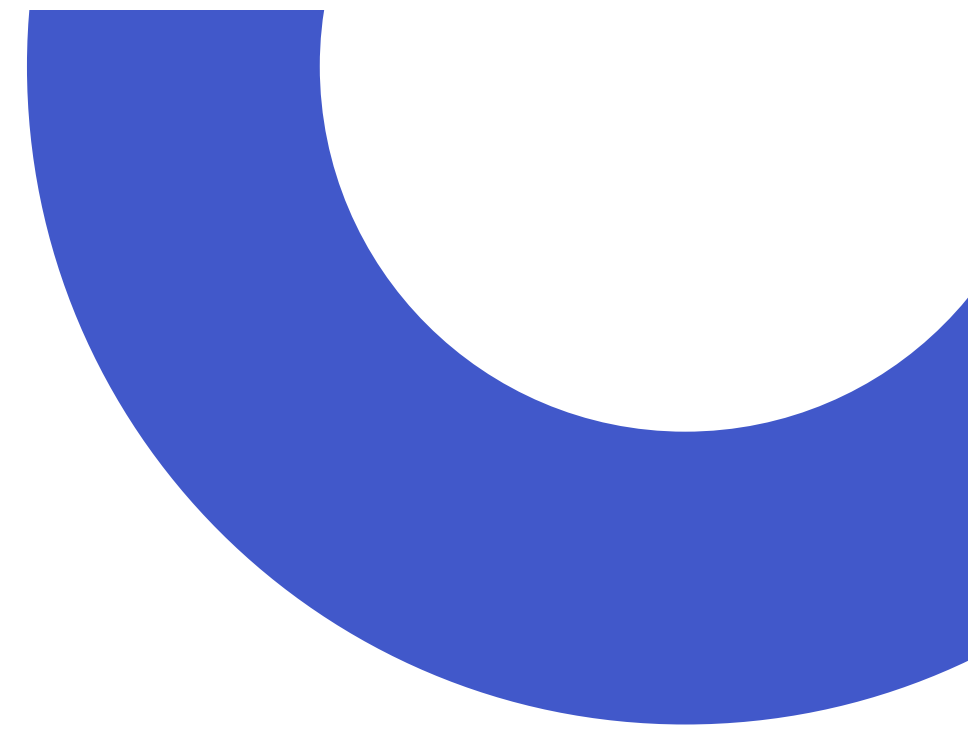
## C. Integración de Bases de Datos Relacionales y NoSQL:

- JSON a CSV/CSV a JSON: Este tipo de transformaciones será necesario para cambiar el formato de los datos entre los distintos sistemas. Se debe utilizar alguna herramienta de integración como Python (pandas, json, csv) para hacer estas conversiones.

## D. Herramientas de Análisis y Análisis de Sentimientos:

- Python: Se utilizará para procesamiento de datos, limpieza, transformación y análisis de sentimientos. Usando bibliotecas como:
  - Pandas para análisis de datos en formato CSV, JSON, etc.
  - NLTK o TextBlob para el análisis de sentimientos de las fuentes de texto.

# Arquitectura de solución.

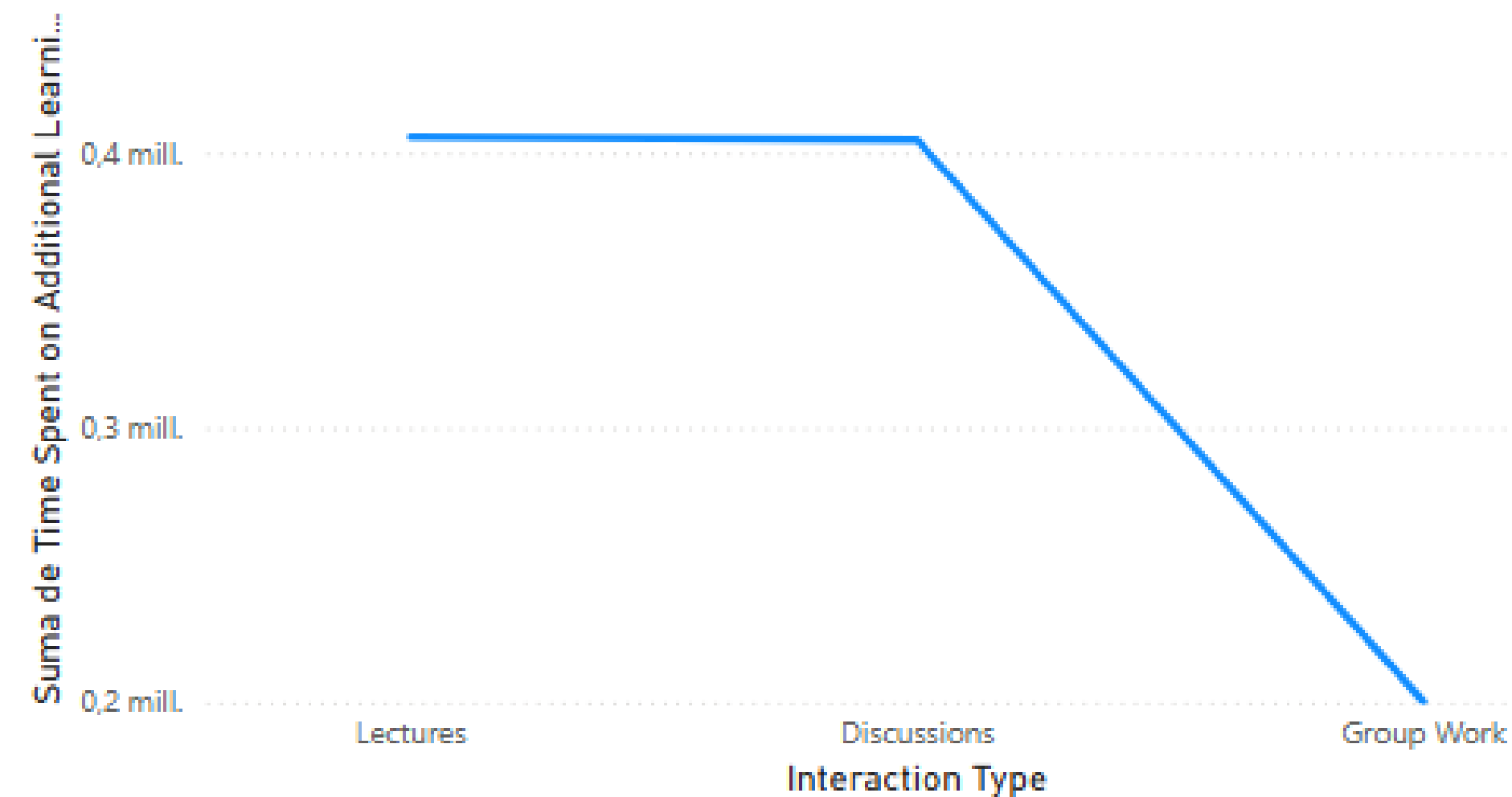


## E. Visualización de Datos:

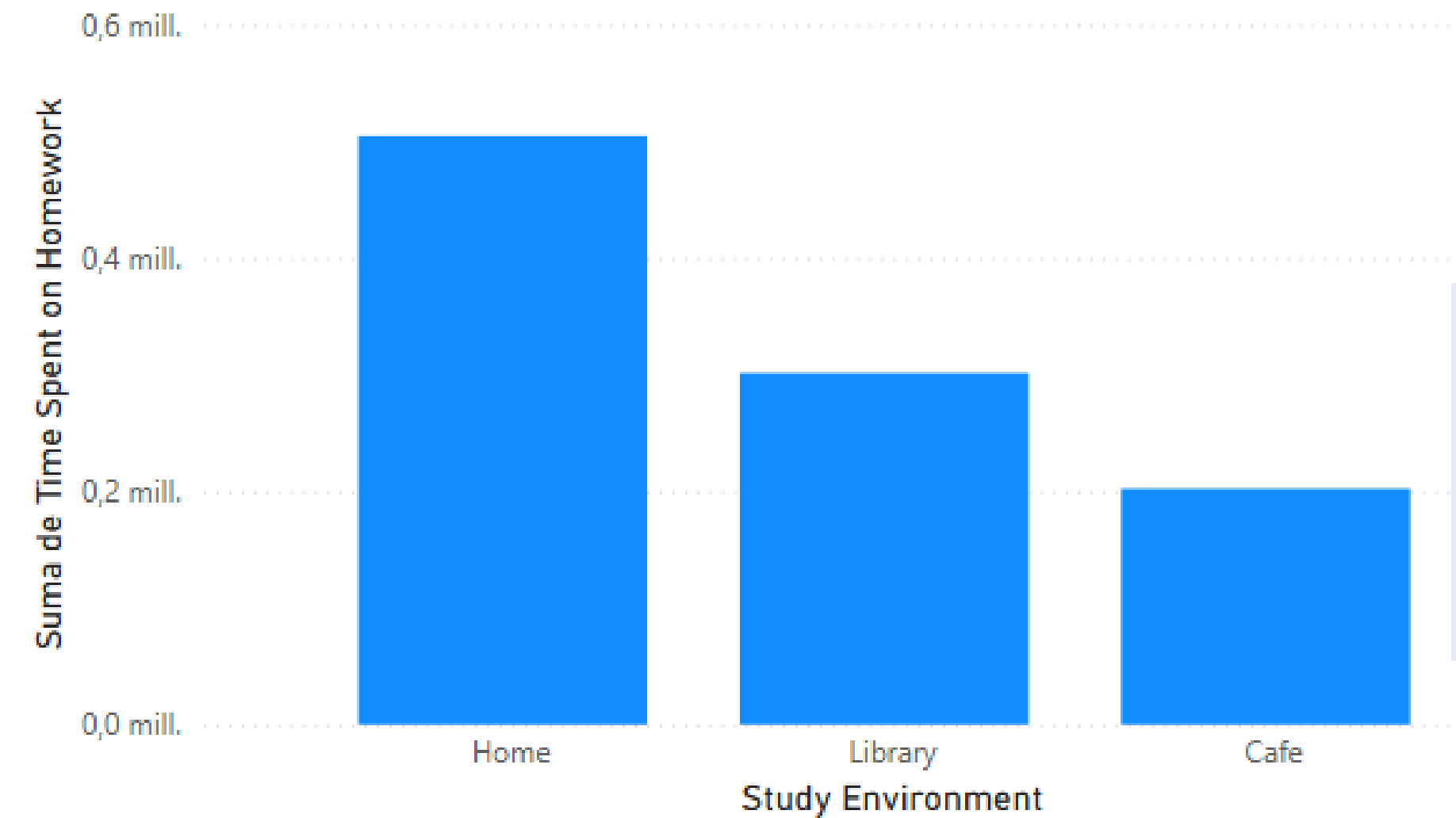
- Power BI : Esta herramienta se utilizarán para crear dashboards interactivos y visualizaciones.
  - Se pueden crear gráficos y tablas para tomar decisiones basadas en la información de eventos deportivos, restaurantes y sitios de esparcimiento, etc.

# Visualización de datos

Suma de Time Spent on Additional Learning por Interaction Type

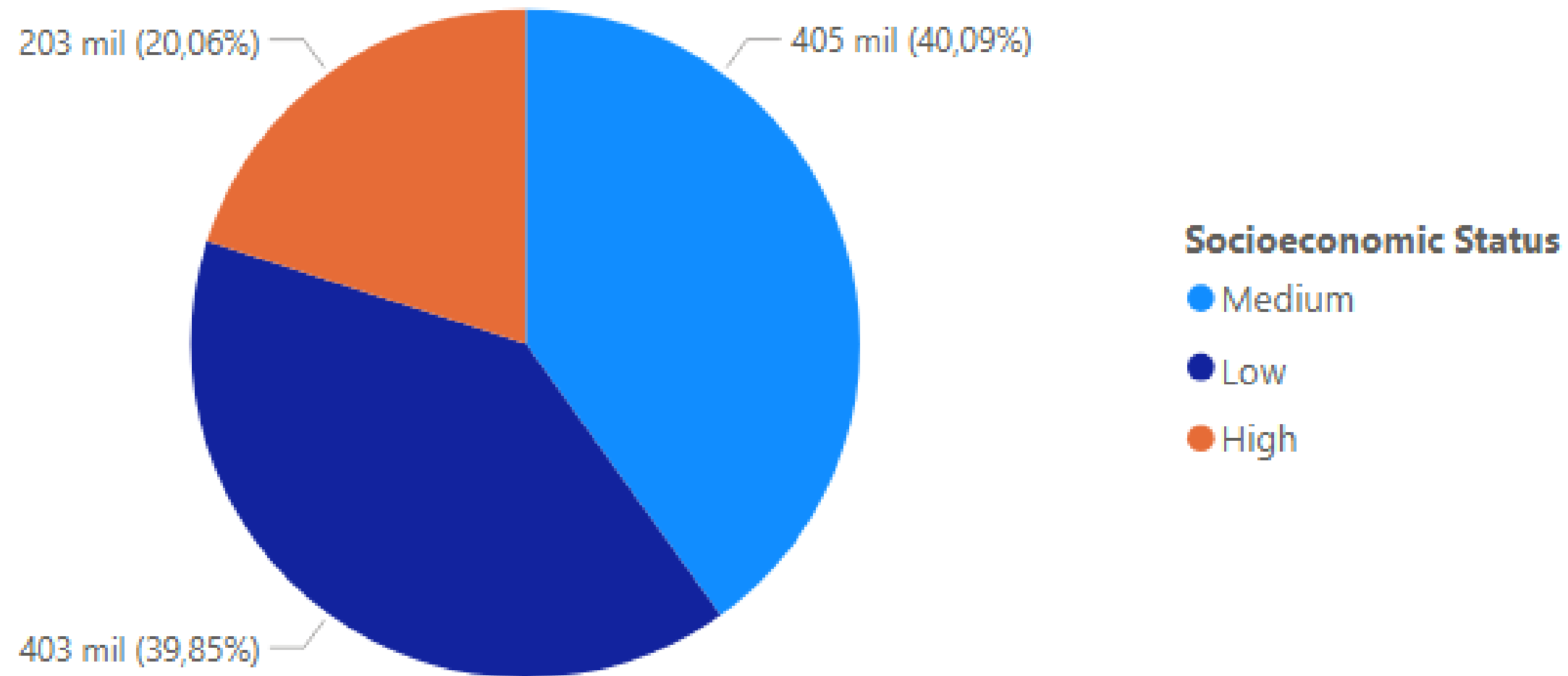


Suma de Time Spent on Homework por Study Environment



# Visualización de datos

Suma de Time Spent on Additional Learning por Socioeconomic Status



**FUENTE:**

[https://app.powerbi.com/view?  
r=eyJrIjoieYTBkNzcwNzAtYzY4Mi00MGMxLWI3YjEtNzQ0Y2E0OTYwNGFlliwidCI6IjY4MmE0ZTZhLWE3N2YtNDk1OC1hM2FjLTllMjY2ZDE4YWEzNyIsImMiOiR9](https://app.powerbi.com/view?r=eyJrIjoieYTBkNzcwNzAtYzY4Mi00MGMxLWI3YjEtNzQ0Y2E0OTYwNGFlliwidCI6IjY4MmE0ZTZhLWE3N2YtNDk1OC1hM2FjLTllMjY2ZDE4YWEzNyIsImMiOiR9)

# Conclusiones y Recomendaciones.

- El uso combinado de bases de datos relacionales (MySQL, SQL Server,) y NoSQL (MongoDB, CouchDB) ha demostrado ser eficaz para manejar diferentes tipos de datos, como datos no estructurados como JSON.
- Los dashboards creados facilitaron la comprensión de datos complejos y los usuarios pueden visualizar las tendencias en tiempo real.
- A medida que el número de registros de un dataset es alto, el tiempo de migración se incrementa, especialmente si no se utilizan herramientas de automatización adecuadas.