



Somos una consultora especializada en análisis y toma de decisiones basada en datos. Nuestro equipo está conformado por profesionales altamente capacitados y con amplia experiencia en proyectos de consultoría de datos, lo que nos permite abordar de manera integral proyectos de diversa complejidad y magnitud.

Nuestra misión es ayudar a nuestros clientes a obtener información valiosa a partir de sus datos para mejorar su toma de decisiones y optimizar sus procesos. Nos destacamos por brindar soluciones personalizadas, adaptadas a las necesidades y objetivos específicos de cada cliente. Trabajamos de manera colaborativa, estableciendo una estrecha relación con nuestros clientes para comprender a fondo su problemática y ofrecerles las mejores soluciones.

Semana 1: Propuesta de proyecto

Entendimiento de la situación actual:

La empresa farmacéutica Pfizer ha delineado su objetivo de desarrollar un nuevo medicamento. Para dar inicio a la planificación de este proyecto, la empresa requiere recopilar datos relevantes que influyan en la esperanza de vida, tasas de fecundidad promedio y otros factores demográficos clave. Esta información será fundamental para determinar la ubicación demográfica más adecuada en la cual invertir y llevar a cabo el desarrollo del medicamento.

Para lograr este objetivo, Pfizer ha contratado a nuestra consultora, con el fin de realizar un estudio sobre la esperanza de vida y otros indicadores relevantes para comprender los factores que influyen en la vitalidad de las personas.

Con el modelo de ML trataremos de resolver estos interrogantes

¿Qué hace que un país tenga mayor esperanza de vida?

¿Cuáles son las características predominantes para que un país tenga mayor esperanza de vida?

Objetivos:

1. Analizar las tendencias actuales de la esperanza de vida en diferentes países y regiones.

2. Identificar los factores determinantes de la esperanza de vida, como las tasas de mortalidad infantil, la calidad de los servicios de atención médica, el acceso a servicios básicos de salud y otros indicadores relevantes.

3. Proporcionar recomendaciones para el desarrollo del nuevo medicamento que promueve la fertilidad y la vitalidad en hombres y mujeres.

Alcance:

Recopilar y analizar datos demográficos, como la esperanza de vida al nacer, la edad media de la población, el índice de envejecimiento y otros, de diferentes fuentes confiables.

El estudio se centrará en países de América del Sur, América del Norte, Centroamérica y Oceanía, utilizando datos recopilados durante los últimos 30 años.

Como posibilidad de continuidad del proyecto, se plantea la expansión a otros países de los otros continentes o la inclusión de factores adicionales.

KPIs:

La metodología de determinación de KPI sigue el concepto de los reconocidos objetivos SMART. El acrónimo S (specific) específico, M (Mesurable) medible, A (Attainable) alcanzable, R (Relevant) relevante, T (Time-bound) con límite de tiempo.

- Esperanza de vida saludable - KPI aumento 0.1% comparando dos últimos años con datos. Es una medida que combina la esperanza de vida al nacer con la proporción de años vividos en buena salud. El índice de esperanza de vida saludable proporciona una visión más completa de la salud de una población al considerar tanto la duración como la calidad de vida.
- Mortalidad Infantil - KPI disminución 0.1% tasa de mortalidad infantil comparando dos últimos años con datos. Es la cantidad de niños que mueren antes de cumplir un año de edad por cada 1,000 nacidos vivos. Es un KPI crítico para evaluar la calidad de los servicios de atención médica materno-infantil y la salud general de la población.
- Esperanza de vida al nacer - KPI aumento 0.1% comparando dos últimos años con datos: Es el indicador principal que mide la esperanza de vida promedio de una población al nacer. Es un KPI fundamental para evaluar la salud y el bienestar general de una población.

Nota: El porcentaje de aumento o disminución está sujeto a modificación, que será probablemente definido la semana entrante.

Repositorio GitHub:

Se creó un repositorio público en GitHub para permitir la colaboración en el desarrollo del proyecto. Se utilizarán diferentes ramas (branches) y se mantendrá un control de versiones del trabajo .[https://github.com/datainsi/Esperanza_de_vida]

Solución propuesta:

1. Análisis de datos: Realizar un análisis exploratorio de los datos disponibles y limpiarlos para su posterior uso.
2. Modelado y análisis: Desarrollar modelos estadísticos y de aprendizaje automático para comprender los factores que influyen en la esperanza de vida.
3. Dashboard de visualización: Crear un dashboard interactivo que presente los resultados del análisis y permita explorar los datos.
4. Informe y recomendaciones: Documentar los hallazgos del análisis y proporcionar recomendaciones basadas en los resultados.

- **Stack tecnológico:**

- Herramientas de trabajo : Visual Studio Code, Jupyter Lab, Google Cloud y Google Colab.
- Lenguaje de programación: Python.
- Bibliotecas: Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn, y Pyspark.
- Herramientas de visualización: Power BI, Streamlit
- Repositorio de código: GitHub

- **Metodología de trabajo:**

Utilizaremos la metodología ágil Scrum para gestionar el desarrollo del proyecto. Realizaremos reuniones periódicas para revisar el progreso, definir tareas y asignar roles. Cada miembro del equipo tendrá responsabilidades específicas y trabajará en colaboración para lograr los objetivos.

- **Diseño detallado - Entregables:**

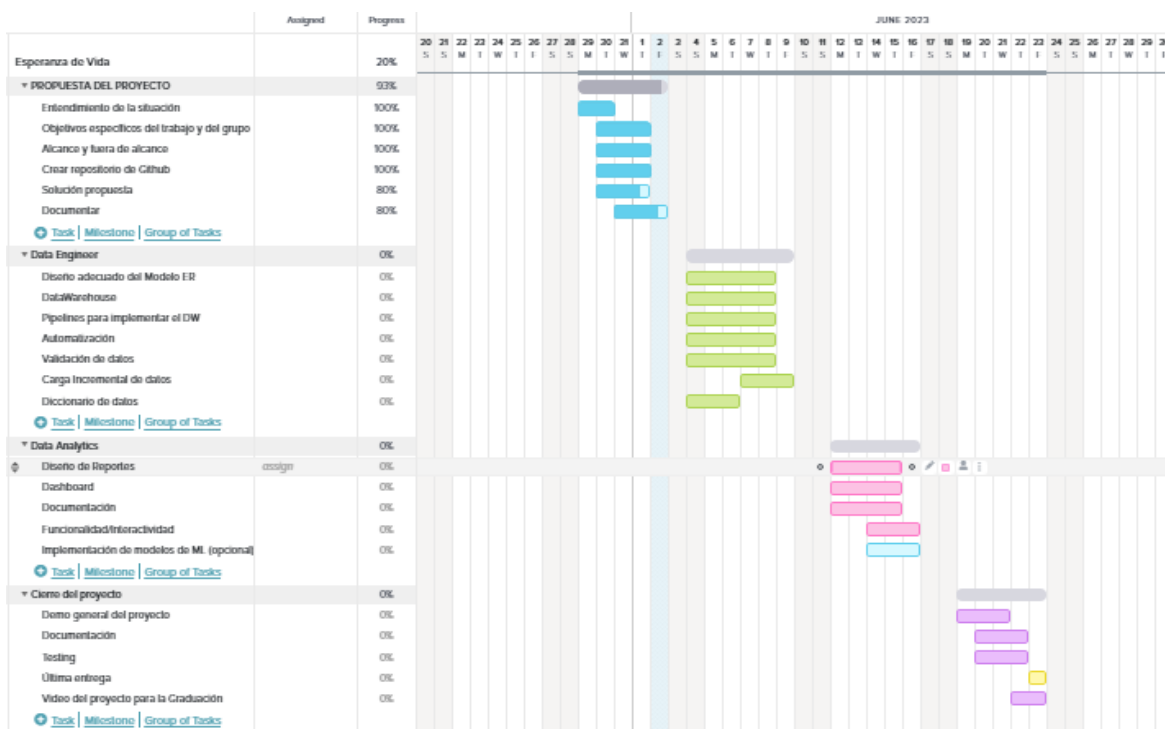
1. Análisis exploratorio de datos: Documento que detalla el proceso de limpieza y exploración de los datos.
2. Modelos estadísticos y de aprendizaje automático: Informe que describe los modelos desarrollados y los resultados obtenidos.
3. Dashboard interactivo: Presentación visual de los resultados del análisis, incluyendo gráficos y visualizaciones interactivas.
4. Informe final y recomendaciones: Documento que resume los hallazgos del análisis y proporciona recomendaciones basadas en los resultados.

- **Equipo de trabajo - Roles y responsabilidades:**

- Data Engineer: Encargado de diseñar e implementar los pipelines de ETL y configurar la infraestructura de datos.
- Data Analyst: Responsable del análisis exploratorio de datos, modelado estadístico y desarrollo de modelos predictivos.
- Data Visualizer: Encargado de crear visualizaciones y el dashboard interactivo para presentar los resultados del análisis.
- Functional Analyst: Asegurar de que el proyecto progrese de manera efectiva y se alcancen los objetivos establecidos al asumir diferentes roles.

- **Cronograma general - Gantt:**

Se creará un diagrama de Gantt que muestre las tareas y los plazos estimados para cada una de las etapas del proyecto. Esto permitirá un seguimiento claro del progreso y asegurará el cumplimiento de los hitos establecidos. [\[https://app.teamgantt.com/projects/gantt?ids=3553277\]](https://app.teamgantt.com/projects/gantt?ids=3553277)



- **Análisis preliminar de calidad de datos:**

Realizaremos un análisis detallado de los datos con los que vamos a trabajar. Analizaremos la descripción de cada columna, los tipos de datos, el método de adquisición y las fechas de adquisición y actualización.

Para la recopilación y el análisis de la data, decidimos basarnos en la base de datos brindada por el Banco Mundial y la Organización Mundial de la Salud (OMS) dado que la confianza y transparencia de su información.