

# ALCANCE DE PROYECTO

Empresa Farmacéutica  
multivitamínico 60+

---

**PROYECTO:** Análisis de la Esperanza de Vida al Nacer para Factibilidad de lanzamiento de multivitamínico a nivel Global.

---

## Contenido Del Documento

<b>ALCANCE DE PROYECTO.....</b>	<b>1</b>
<b>Empresa Farmacéutica .....</b>	<b>1</b>
<b>1. Introducción.....</b>	<b>3</b>
<b>2. Contexto del Alcance .....</b>	<b>4</b>
<b>3. Descripción del Alcance.....</b>	<b>5</b>
<b>3.1 Objetivos del Proyecto .....</b>	<b>5</b>
<b>3.2 Alcance del Proyecto .....</b>	<b>6</b>
<b>3.3 Requisitos del Cliente .....</b>	<b>6</b>
<b>4. Entregables .....</b>	<b>8</b>
<b>5. Cronograma y Fechas Clave.....</b>	<b>9</b>
<b>4.1 SPRINT # 1 .....</b>	<b>9</b>
<b>4.2 SPRINT # 2 .....</b>	<b>9</b>
<b>4.3 SPRINT # 3 .....</b>	<b>9</b>
<b>6. Recursos y Personal .....</b>	<b>10</b>
<b>7. Riesgos y Mitigaciones.....</b>	<b>10</b>

## 1. Introducción

Este documento presenta el alcance del proyecto de consultoría de datos, cuyo propósito es analizar una serie de factores que influyen en la decisión de lanzar o no un producto farmacéutico en mercados específicos o regiones. Nuestro cliente, una destacada empresa en la industria farmacéutica, busca comprender estos factores y su variabilidad regional con el objetivo de evaluar la viabilidad del lanzamiento de un nuevo producto: un multivitamínico destinado a la población de edad avanzada (60+).

## 2. Contexto del Alcance

En la industria farmacéutica, una serie de factores críticos influyen en la toma de decisiones estratégicas. Estos factores abarcan aspectos socioeconómicos, culturales y sociales que desempeñan un papel fundamental en la evaluación de oportunidades de negocio en distintas regiones del mundo. Uno de estos factores es la esperanza de vida al nacer, que se define como la estimación promedio de años que una persona puede esperar vivir al nacer en una región o país específico. La esperanza de vida al nacer refleja la calidad de vida, las condiciones de salud y la longevidad de la población, y sirve como un indicador esencial para la demanda y necesidad de productos y servicios médicos.

Sin embargo, la toma de decisiones estratégicas en la industria farmacéutica no se limita únicamente a la esperanza de vida al nacer. También se consideran factores socioeconómicos, culturales y sociales que incluyen la distribución de población en áreas urbanas y rurales, el PBI per cápita, el nivel educacional, la inflación, el crecimiento poblacional y la accesibilidad a los servicios de atención médica (entre otros). Estos elementos en conjunto proporcionan información valiosa sobre las oportunidades de negocio en diferentes regiones del mundo.

Las empresas farmacéuticas evalúan una combinación de estos factores para identificar áreas geográficas con alto potencial de mercado y demanda de productos médicos. Este contexto subraya la importancia de considerar una variedad de factores, incluida la esperanza de vida al nacer, en la búsqueda de oportunidades de negocio en la industria farmacéutica.

### 3. Descripción del Alcance

#### 3.1 Objetivos del Proyecto

Realizar un análisis exhaustivo del impacto de los factores mencionados anteriormente en los países seleccionados, con el propósito de proporcionar a una empresa farmacéutica información crítica para la toma de decisiones estratégicas relacionadas con el lanzamiento de un nuevo producto en aquellos países donde sea más probable su rentabilidad.

Los objetivos específicos de este proyecto son los siguientes:

1. **Evaluar la Esperanza de Vida y Demanda Potencial:** Analizar la esperanza de vida en las regiones seleccionadas como indicador clave para determinar la viabilidad y la demanda potencial del nuevo multivitamínico entre la población de 60 años o más.
2. **Segmentar Audiencia Potencial por Áreas Urbanas y Rurales:** Examinar la distribución de la población en áreas urbanas y rurales para identificar y segmentar adecuadamente el público objetivo en cada región, considerando las diferencias de estilo de vida y necesidades.
3. **Evaluar Factores Socioeconómicos y Culturales:** Analizar indicadores socioeconómicos y culturales, como el ingreso per cápita y el nivel de educación en las regiones estudiadas, para comprender su impacto en la disposición de la población mayor de edad a invertir en su salud y bienestar.
4. **Identificar Regiones con Mayor Proporción de Población de 60+:** Examinar los rangos de edades y sus proporciones en cada región con el fin de identificar aquellas áreas con una concentración significativa de personas mayores de 60 años.
5. **Evaluar Factores Económicos:** Determinar los valores de inflación y el crecimiento del Producto Interno Bruto (PIB) en cada región, ya que estos factores económicos influyen en el comportamiento de consumo de productos de salud y bienestar.
6. **Analizar Crecimiento Poblacional:** Evaluar el crecimiento poblacional en cada región, ya que las regiones con un mayor crecimiento pueden albergar un número creciente de potenciales clientes en el segmento de edad de 60+.
7. **Evaluar la Accesibilidad a Servicios de Atención Médica:** Evaluar la disponibilidad de personal de salud, incluyendo médicos y profesionales de la atención médica, en cada región seleccionada. Este análisis ayudará a comprender la accesibilidad de la población a servicios de atención médica y de salud, lo cual puede influir en la necesidad de suplementos vitamínicos en personas mayores.

## 3.2 Alcance del Proyecto

El alcance total del proyecto contempla:

- Recopilar y analizar datos de al menos 10 factores o bases de datos del Banco Mundial de Datos, utilizando su API. Se permitirá la obtención de datos de fuentes complementarias si contribuyen significativamente a la toma de decisiones.
- Realizar un análisis exploratorio de datos (EDA) y un proceso de extracción, transformación y carga de datos (ETL) exhaustivo para normalizar los datos y persistir solo aquellos relevantes para los objetivos planteados.
- Definir y medir al menos 5 KPIs (Indicadores Clave de Rendimiento) relevantes para este contexto y proyecto.
- Abarcar 44 países distribuidos en los 5 continentes (América, Europa, África, Asia y Oceanía), con un historial mínimo de 35 años de datos disponibles.
- Construir un modelo de Machine Learning que evalúe si un país o región es rentable o no para el lanzamiento del nuevo producto, basado en los datos recopilados y los KPIs definidos.
- Persistir los datos en la Nube utilizando Google Cloud Platform (GCP) y generar un modelo relacional para consulta por parte de áreas técnicas.
- Crear un Pipeline de Datos automatizado para realizar cargas incrementales de datos en GCP.
- Construir un Dashboard Interactivo con una visión estratégica para la toma de decisiones sobre el lanzamiento de productos, centrándose en el multivitamínico.

## 3.3 Requisitos del Cliente

El cliente, una empresa farmacéutica líder en la industria, presenta los siguientes requisitos clave para el proyecto:

1. **Análisis Exhaustivo de la Esperanza de Vida:** El cliente busca un análisis completo y en profundidad de la esperanza de vida al nacer en diferentes regiones del mundo. Esto incluye la recopilación de datos históricos y actuales, así como la identificación de tendencias a lo largo del tiempo.
2. **Factores Determinantes:** Además de la esperanza de vida en sí, el cliente está interesado en identificar y comprender los factores determinantes que influyen en la esperanza de vida. Estos factores pueden incluir aspectos culturales, hábitos humanos, acceso a la atención médica, condiciones socioeconómicas y otros que puedan contribuir a la variabilidad en la esperanza de vida.
3. **Valor para la Toma de Decisiones:** El análisis debe proporcionar información valiosa que respalde la toma de decisiones de la empresa farmacéutica. Esto incluye la evaluación de oportunidades de negocio en diferentes regiones y la identificación de mercados donde podría ser rentable lanzar nuevos productos, como el multivitamínico para personas mayores de edad.
4. **Precisión y Confiabilidad:** Se requiere un alto nivel de precisión y confiabilidad en el análisis de datos y la metodología utilizada. Los resultados deben ser basados en datos sólidos y metodologías estadísticas adecuadas.

5. **Presentación de Resultados:** El cliente espera una presentación clara y efectiva de los resultados del análisis, incluyendo visualizaciones gráficas, informes detallados y resúmenes ejecutivos que sean fácilmente comprensibles para los equipos de toma de decisiones.
6. **Cumplimiento de Plazos:** El proyecto debe cumplir con los plazos acordados y mantener un flujo de trabajo eficiente para garantizar la disponibilidad oportuna de resultados.
7. **Protección de Datos:** La empresa farmacéutica enfatiza la importancia de la privacidad y la seguridad de los datos recopilados y procesados. Se deben implementar medidas adecuadas para proteger la confidencialidad de la información.

Además, el cliente solicitó la medición de los siguientes **KPIs**:

1. **Aumento de la Esperanza de Vida en un 1% anual**  
Objetivo: Evaluar anualmente el incremento en la esperanza de vida en las regiones seleccionadas.
2. **Aumento de la Población Mayor a 65 años en un 2% anual**  
Objetivo: Evaluar el crecimiento anual de la población de 65 años o más en las regiones estudiadas.
3. **Aumento de la Población Urbana en un 1% anual**  
Objetivo: Evaluar anualmente el aumento de la población en áreas urbanas en las regiones seleccionadas.
4. **Aumento de la Población con Educación Secundaria Completa en un 2% anual**  
Objetivo: Evaluar anualmente el incremento de la población con educación secundaria completa en las regiones estudiadas.
5. **Aumento del Producto Interno Bruto (PBI) en un 1% anual**  
Objetivo: Evaluar anualmente el crecimiento del Producto Interno Bruto (PBI) en las regiones seleccionadas.
6. **Disminución de la Inflación en un 1% anual**  
Objetivo: Evaluar anualmente la reducción de la tasa de inflación en las regiones estudiadas.
7. **Aumento del Gasto en Salud en un 1% anual**  
Objetivo: Evaluar anualmente el gasto de Salud por parte de la población en las regiones estudiadas.
8. **Eficacia del Modelo de Machine Learning superior al 80%**  
Objetivo: Mantener una eficacia superior al 80% en el modelo de Machine Learning, diseñado para evaluar la factibilidad del lanzamiento del producto en una determinada región.

## 4. Entregables

Los entregables del proyecto incluirán:

- **Documentación**
  - ✓ Documento de alcance y diagrama de Gantt: Un documento que describe el alcance completo del proyecto y un diagrama de Gantt que muestra la planificación temporal de las tareas.
  - ✓ Stack elegido y fundamentación: Documentación que describe detalladamente las tecnologías, herramientas y plataformas seleccionadas para el proyecto, junto con la justificación de su elección.
  - ✓ Flujo de trabajo: Una descripción detallada del flujo de trabajo utilizado para la recopilación, transformación y carga de datos, incluyendo todos los pasos y tareas involucradas.
  - ✓ Modelo de Entidad Relación de la Base de Datos a persistir en Google Cloud Platform (GCP): Un modelo que representa la estructura de la base de datos relacional utilizada en GCP, incluyendo tablas, relaciones y atributos.
  - ✓ Reporte realizado en el proceso de Data Engineering: Un informe que documenta todo el proceso de ingeniería de datos, incluyendo la limpieza, transformación y preparación de los datos.
- **Dashboard interactivo** y visualmente atractivo que presenta los resultados del análisis de datos y los KPIs relevantes. Debe ser fácil de entender y utilizar, diseñado para respaldar la toma de decisiones estratégicas en el lanzamiento de productos.
- **Modelo de Machine Learning** que pueda evaluar si un país o región es rentable o no para el lanzamiento del nuevo producto, basado en los datos recopilados. Esto debe incluir detalles sobre el proceso de construcción, las métricas de evaluación y la explicación del modelo.
- Todos los códigos, scripts y pipelines utilizados en el proyecto, alojados en un **repositorio GitHub** u otra plataforma de control de versiones. Esto incluye tanto el código para la recopilación y procesamiento de datos como el código del modelo de machine learning.



## 5. Cronograma y Fechas Clave

El proyecto será llevado a cabo utilizando la metodología ágil **SCRUM**. Será dividido en 3 SPRINTs, donde la duración de cada uno será de 2 semanas. A continuación, se presenta una vista general de las fechas clave:

### 4.1 SPRINT # 1

- Fecha Inicio: 24/10/2023
- Fecha Fin: 03/11/2023
- Contemplara las siguientes tareas principales:
  - ✓ Identificación de Factores relevantes
  - ✓ Extracción de datos y almacenaje en Cloud
  - ✓ EDA y análisis preliminar de calidad de los datos
  - ✓ Implementación stack tecnológico detallado
  - ✓ Metodología de trabajo, roles y responsabilidades
  - ✓ Cronograma general- Gantt
  - ✓ DEMO con PO. ➔ 1/11/2023.

### 4.2 SPRINT # 2

- Fecha Inicio: 06/11/2023
- Fecha Fin: 18/11/2023
- Contemplara las siguientes tareas principales:
  - ✓ ETL Definitivo
  - ✓ Persistencia de Datos en Cloud
  - ✓ Pipeline en GCP automatizado
  - ✓ DER de la BBDD
  - ✓ Diccionario de Datos
  - ✓ documentación detallada de la etapa de Data Engineering
  - ✓ EDA y análisis preliminar de calidad de los datos
  - ✓ MVP de Producto de ML/Dashboard
  - ✓ DEMO con PO. ➔ 15/11/2023.

### 4.3 SPRINT # 3

- Fecha Inicio: 20/11/2023
- Fecha Fin: 02/12/2023
- Contemplara las siguientes tareas principales:
  - ✓ DashBoard Interactivo
  - ✓ Modelo de Machine Learning Productivo
  - ✓ documentación del modelo de ML elegido
  - ✓ Video Explicativo del Proyecto
  - ✓ DEMO con PO. ➔ 29/11/2023.

## 6. Recursos y Personal

El equipo de desarrollo estará formado por 5 miembros, cada uno con roles y responsabilidades específicas. Los detalles del equipo se podrán encontrar en el documento de Roles y Responsabilidades.

## 7. Riesgos y Mitigaciones

Se han identificado los siguientes riesgos:

1. **Riesgo: Datos Inconsistentes o de Mala Calidad**
  - **Mitigación:** Realizar una exhaustiva limpieza de datos durante el proceso de ETL, implementar validaciones de integridad y aplicar transformaciones necesarias para abordar la inconsistencia de datos.
2. **Riesgo: Retrasos en la Recopilación de Datos**
  - **Mitigación:** Establecer un plan de contingencia para tratar retrasos en la obtención de datos, como la posibilidad de utilizar datos de respaldo o ajustar el calendario del proyecto.
3. **Riesgo: Problemas de Privacidad de Datos**
  - **Mitigación:** Implementar medidas de seguridad y privacidad de datos, como enmascaramiento de información sensible y cumplimiento con regulaciones de privacidad, para proteger la información recopilada.
4. **Riesgo: Escalabilidad de Recursos en la Nube**
  - **Mitigación:** Monitorear de cerca la escalabilidad de recursos en la nube y ajustar los recursos según sea necesario para garantizar un rendimiento óptimo y cumplir con los plazos.
5. **Riesgo: Complejidad del Modelo de Machine Learning**
  - **Mitigación:** Realizar pruebas y validaciones exhaustivas del modelo de machine learning para asegurar que sea preciso y adecuado para los objetivos.
6. **Riesgo: Cambios en los Requisitos del Cliente**
  - **Mitigación:** Establecer una comunicación constante con el cliente y documentar los requisitos de manera detallada.
7. **Riesgo: Fallos en la Automatización del Pipeline de Datos**
  - **Mitigación:** Realizar pruebas exhaustivas del pipeline de datos y establecer un proceso de monitoreo y alertas para detectar y abordar posibles fallos de automatización de manera oportuna.