



Reporte del Proyecto Fianl: Expectativa de vida al nacer

ML: Gonzalo Ducca <gonzaloducca@gmail.com>

DA: Juan P. Bertone <bertonejpb@gmail.com>

DE: Juan E. Flórez-Coronel <juan.florez@upr.edu>

DA: Valentino Caputa <caputavalentino@gmail.com>

ML: Carlos Madoery <ccmadoery@gmail.com>

Product Owner:

Melina Griffio <meligriffio@gmail.com>

Henry Mentor:

Pía Ruiz <mpiaruiz@gmail.com>

HENRY Data Science Part Time 3

Revision 1.0

November 15, 2023

Table 1. Document Revision History

Revision	Date	Reason for Revision Change(s)
0.0	10-20-23	Creación del documento
1.0	11-01-23	Documentación Sprint 1
2.0	11-15-23	Documentación Sprint 2

1 Planteamiento del problema

Consultamos bases de datos del Banco Mundial y decidimos utilizar como referencia temas clave que proporcionan indicadores que influyen en la esperanza de vida de los habitantes de un país. Estos temas seleccionados son la educación, la salud, la economía, el desarrollo de la ciencia y la tecnología y el ámbito social. Planeamos recopilar indicadores de cada tema, crear las bases de datos con las que queremos trabajar y luego vincular estas variables para establecer relaciones. Visualizamos este proyecto como un producto que se puede ofrecer a una empresa que busque invertir en ciencia y desarrollo. Una vez completado todo el trabajo de análisis de datos, crearemos modelos de aprendizaje automático especificando ciertos parámetros, como el porcentaje del invertido en educación o ciencia. Estos modelos luego nos proporcionarán una esperanza de vida estimada para un país en particular.

1.1 Objetivos

Analizar la expectativa de vida en 30 países en base a 5 factores:

- Educación
- Salud
- Economía
- Ciencia y Tecnología
- Social

1.2 Público objetivo

El público objetivo de este proyecto incluye entidades públicas y privadas con intereses creados en el tema y usuarios cotidianos que quieran conocer la esperanza de vida esperada en un país determinado. Las instituciones públicas, como las agencias gubernamentales, pueden utilizar los conocimientos y modelos generados para informar decisiones políticas y asignar recursos de manera efectiva en áreas relacionadas con la educación, la atención médica y la investigación científica. Por otro lado, las empresas privadas, especialmente aquellas involucradas en industrias relacionadas con la ciencia y la tecnología, pueden beneficiarse de los hallazgos del proyecto al tomar decisiones de inversión e iniciativas de responsabilidad social corporativa. Este proyecto está dirigido a una amplia gama de organizaciones que buscan tomar decisiones informadas que impacten el bienestar y el desarrollo de los países.

1.3 Alcance del proyecto

El alcance del proyecto implica un análisis exhaustivo de la esperanza de vida en 30 países de todo el mundo durante los últimos 35 años. Este análisis abarcará cinco temas clave dentro

de cada uno de los cinco factores influyentes, a saber, educación, salud, economía, ciencia y tecnología y ámbito social. El objetivo es proporcionar una visión detallada y basada en datos de los factores que afectan la esperanza de vida en estos países durante el período de tiempo especificado. El producto final incluirá un servicio web donde se pueda acceder a la información brindada por el análisis y un dashboard interactivo con insights claves para la interpretación de los datos.

2 Obtención de los datos

La información fue obtenida utilizando la siguiente API World Bank API.

2.1 World Bank API

La información de la API de World Bank tiene los siguientes tópicos:

id	value
1	Agriculture & Rural Development
2	Aid Effectiveness
3	Economy & Growth
4	Education
5	Energy & Mining
6	Environment
7	Financial Sector
8	Health
9	Infrastructure
10	Social Protection & Labor
11	Poverty
12	Private Sector
13	Public Sector
14	Science & Technology
15	Social Development
16	Urban Development
17	Gender
18	Millenium development goals
19	Climate Change
20	External Debt
21	Trade

Figure 1. Tópicos World Bank

Y las siguientes economías:

Cada tópico tiene factores asociados:

	id	value	aggregate	longitude	latitude	region	adminregion	lendingType	incomeLevel	capitalCity
0	ABW	Aruba	False	-70.0167	12.51670	LCN		LNK	HIC	Oranjestad
1	AFE	Africa Eastern and Southern	True	NaN	NaN					
2	AFG	Afghanistan	False	69.1761	34.52280	SAS	SAS	IDX	LIC	Kabul
3	AFW	Africa Western and Central	True	NaN	NaN					
4	AGO	Angola	False	13.2420	-8.81155	SSF	SSA	IBD	LMC	Luanda
...
261	XKX	Kosovo	False	20.9260	42.56500	ECS	ECA	IDX	UMC	Pristina
262	YEM	Yemen, Rep.	False	44.2075	15.35200	MEA	MNA	IDX	LIC	Sana'a
263	ZAF	South Africa	False	28.1871	-25.74600	SSF	SSA	IBD	UMC	Pretoria
264	ZMB	Zambia	False	28.2937	-15.39820	SSF	SSA	IDX	LMC	Lusaka
265	ZWE	Zimbabwe	False	31.0672	-17.83120	SSF	SSA	IDB	LMC	Harare

Figure 2. Economías World Bank

	id	value
0	BM.GSR.ROYL.CD	Charges for the use of intellectual property, ...
1	BX.GSR.ROYL.CD	Charges for the use of intellectual property, ...
2	TX.VAL.TECH.MF.ZS	High-technology exports (% of manufactured exp...
3	TX.VAL.TECH.CD	High-technology exports (current US\$)
4	IP.PAT.NRES	Patent applications, nonresidents
5	IP.PAT.RESD	Patent applications, residents
6	GB.XPD.RSDV.GD.ZS	Research and development expenditure (% of GDP)
7	SP.POP.SCIE.RD.P6	Researchers in R&D (per million people)
8	IP.JRN.ARTC.SC	Scientific and technical journal articles
9	SP.POP.TECH.RD.P6	Technicians in R&D (per million people)

Figure 3. Factores Ciencia y Tecnología World Bank

2.2Tópicos Posibles

Contando valores faltantes:

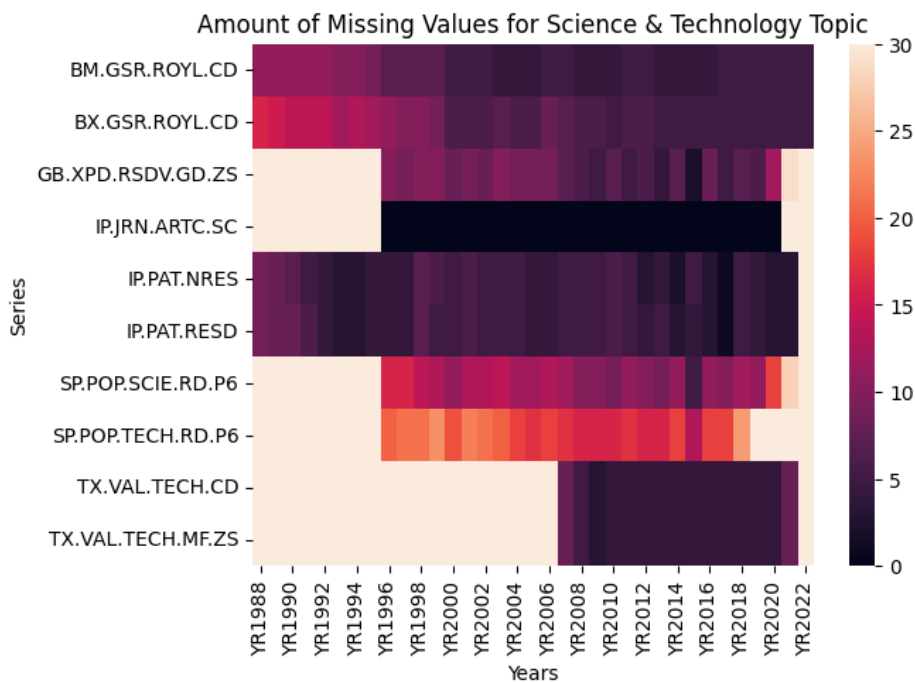


Figure 4. Análisis de valores Nan

3 Desarrollo del Proyecto

3.1 Roles del Equipo

Table 2. Roles del Equipo

Role	Name
Machine Learning	Gonzalo Ducca
Machine Learning	Carlos Madoery
Data Analytics	Valentino Caputa
Data Analytics	Juan P. Bertone
Data Engineer	Juan E. Flórez-Coronel

3.2 Diagrama de Gaant

Esperanza de vida

Sprint 1

Nombre	Responsable	Fecha	Estado	Cronograma - Start	Cronograma - End
KPI	caputavalentino@gmail.com	2023-11-01	Listo	2023-10-23	2023-11-01
Stack tecnológico	Juan Pablo Bertone	2023-10-27	Listo	2023-10-23	2023-11-01
Documentación	Gonzalo Ducca, Carlos Madoery	2023-10-28	En curso	2023-10-23	2023-11-01
EDA preliminar	Juan E Florez-Coronel	2023-10-31	Listo	2023-10-23	2023-11-01
Github	Juan E Florez-Coronel	2023-10-27	Listo	2023-10-23	2023-11-01
Desde el 2023-10-27 hasta el 2023-11-01				2023-10-23	2023-11-01

Sprint 2

Nombre	Responsable	Fecha	Estado	Cronograma - Start	Cronograma - End
DW automatizado	Juan E Florez-Coronel	2023-11-07	No iniciado	2023-11-01	2023-11-08
Tablas	Juan E Florez-Coronel, Gonzalo Ducca		No iniciado	2023-11-01	2023-11-08
Carga incremental	Carlos Madoery	2023-11-10	No iniciado	2023-11-08	2023-11-15
Big Data y servicios cloud.	Juan Pablo Bertone	2023-11-13	No iniciado	2023-11-08	2023-11-15
Desde el 2023-11-07 hasta el 2023-11-13				2023-11-01	2023-11-15

Sprint 3

Nombre	Responsable	Fecha	Estado	Cronograma - Start	Cronograma - End
Storytelling	caputavalentino@gmail.com, Juan Pablo Bertone	2023-11-28	No iniciado	2023-11-15	2023-11-29
Dashboard y reportes	valentino@gmail.com, Juan E Florez-Coronel, Juan Pablo Bertone	2023-11-25	No iniciado	2023-11-15	2023-11-29
Recomendaciones	Carlos Madoery	2023-11-29	No iniciado	2023-11-15	2023-11-29
KPIs y datos hallados	Carlos Madoery	2023-11-27	No iniciado	2023-11-15	2023-11-29
Retomar hitos	Gonzalo Ducca	2023-11-25	No iniciado	2023-11-15	2023-11-29
Modelos ML	Gonzalo Ducca, Carlos Madoery	2023-11-21	No iniciado	2023-11-15	2023-11-29
Reporte Vis. geografica	Juan Pablo Bertone	2023-11-23	No iniciado	2023-11-15	2023-11-29
Desde el 2023-11-21 hasta el 2023-11-29				2023-11-15	2023-11-29

Figure 5. Diagrama de Gantt

CConclusiones y Lecciones Aprendidas

bla