FLUJOS MIGRATORIOS EN ARGENTINA (1960 – 2022)



Sprint 1: Entendimiento de situación, alcance de propuesta y solución tecnológica para implementación de pipeline.

**I. Entendimiento de la situación propuesta**

Los flujos migratorios, el movimiento de personas de un lugar a otro, son causados ​​por una compleja interacción de varios factores. Estos factores se pueden clasificar en términos generales en factores de empuje y factores de atracción. Es importante tener en cuenta que la migración es un fenómeno multifacético, y las decisiones individuales de migrar están influenciadas por una combinación de estos factores en lugar de uno solo. Aquí hay una descripción general de las principales causas de los flujos migratorios:

|  |  |
| --- | --- |
| **Factores de presión** | **Factores de atracción** |
| **Razones económicas:** La falta de oportunidades laborales, los salarios bajos y las malas condiciones económicas en el país de origen pueden empujar a las personas a buscar mejores perspectivas económicas en otros lugares. | **Mejores oportunidades económicas:** la promesa de salarios más altos, mejores perspectivas laborales y mejores niveles de vida en un país de destino puede atraer a inmigrantes que buscan un progreso económico. |
| **Conflicto y violencia:** los conflictos armados, las guerras civiles, la inestabilidad política y la violencia pueden obligar a las personas a huir de sus hogares en busca de seguridad y protección. | **Educación y capacitación:** el acceso a instalaciones de educación y capacitación de calidad puede atraer a inmigrantes que desean mejorar sus habilidades y conocimientos. |
| **Factores ambientales:** los desastres naturales como huracanes, terremotos, sequías y otros desafíos ambientales pueden hacer que las áreas sean inhabitables, lo que lleva a las personas a migrar a regiones más seguras. | **Estabilidad política y seguridad:** los países con sistemas políticos estables, estado de derecho y bajas tasas de criminalidad pueden ser atractivos para las personas que buscan seguridad y protección. |
| **Falta de servicios:** el acceso limitado a la atención médica, la educación y otros servicios esenciales puede llevar a las personas y las familias a buscar mejores condiciones de vida en otros lugares. | **Reunificación familiar:** el deseo de reunirse con miembros de la familia que ya han migrado puede ser un fuerte factor de atracción para muchos migrantes. |
| **Factores políticos:** los regímenes represivos, la falta de libertades políticas y los abusos de los derechos humanos pueden motivar a las personas a abandonar sus países de origen en busca de asilo político. | **Vínculos culturales y sociales:** las comunidades de inmigrantes de la misma región o antecedentes culturales pueden crear un entorno de apoyo que aliente a otros a unirse a ellos. |
|  | **Mejor calidad de vida:** el acceso a mejores servicios de salud, servicios sociales, infraestructura y calidad de vida en general puede ser atractivo para los migrantes que buscan un mejor nivel de vida. |
|  | **Políticas de inmigración abiertas:** los países con políticas de inmigración más indulgentes y actitudes acogedoras hacia los inmigrantes pueden atraer a personas que buscan un nuevo hogar. |
|  | **Turismo y Trabajo Temporal:** Algunos migrantes se mudan temporalmente por trabajo de temporada o turismo, pero pueden decidir quedarse más tiempo o permanentemente debido a las experiencias positivas que tienen. |

Fuente: Simpson, Nicole B. Demographic and economic determinants of migration. IZA World of Labor. https://wol.iza.org/articles/demographic-and-economic-determinants-of-migration/long.

The push-pull factors of migration-geography.Tutor2u. https://www.tutor2u.net/geography/reference/the-push-pull-factors-of-migration.

Kyaing Kyaing Thet. Pull and Push Factors of Migration: A Case Study in the Urban Area of Monywa Township, Myanmar. World of Statistics. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.worldofstatistics.org/files/2014/03/Pull-and-Push-Factors-of-Migration-Thet.pdf.

Es importante reconocer que estos factores están interconectados y pueden influir en los flujos migratorios de diferentes maneras. Además, la percepción de estos factores puede variar de persona a persona, dando lugar a una amplia gama de experiencias y motivaciones migratorias.

**II. Objetivos: específicos del trabajo y del grupo**

**II.1 General:**

Implementar un pipeline que permita explorar, analizar y predecir los flujos migratorios de los próximos años en Argentina con el objetivo de hacer un proyecto en donde el gobierno del país tenga información con la cual tomar decisiones en un futuro cercano.

**II.2 Específicos del trabajo**

Diseñar un pipeline que permita monitorear el ciclo completo de los datos de flujos de migración en Argentina del año 1960 al 2021.

Desarrollar procesos automatizados en carga incremental de datos y procesamientos intermedios para el análisis estadístico, visualizaciones y machine learning.

Producir una base de datos para realizar análisis estadístico, visualizaciones y machine Learning, y automatizar su carga.

Comprender el comportamiento o mecánica de los flujos migratorios.

Brindar KPIs útiles para entender el fenómeno y tener una idea de control para tomar decisiones.

**II.3 Específicos del grupo**

Aprender desde 0 tecnologías para implementar y diseñar pipelines.

Fortalecer las habilidades técnicas de automatización de subprocesos en pipeline.

Fortalecer las habilidades blandas y de comunicación bajo presión.

Afianzar y perfeccionar conocimientos aprendidos en la carrera de Data Science.

**III. Alcance**

Diseñar un pipeline usando tecnología cloud computing que permita automatizar el ciclo de dato de los flujos migratorios en argentina del año 1960 al 2022 para explicar el fenómeno de desplazamiento de personas, comprender, crear métricas de monitoreo y KPIs que permitan tomar decisiones a un gobierno.

Acciones a realizar:

* Carga inicial en ambiente cloud.
* Automatización de carga incremental y limpieza de datos en base de datos Google Cloud Storage.
* Automatización del envío de datos desde nuestro Datalake hacía nuestro Datawherehouse en BigQuery.
* Análisis de datos (en ambiente cloud).
* Machine learning (en ambiente cloud).
* Visualización de datos (en ambiente cloud).

**IV. Almacenamiento de código**

Repositorio de Github: <https://github.com/RomanAichino/PF_Migraciones_Grupo_2>

**V. Solución propuesta:**

La solución esta diseñada para ser implementada usando como principal plataforma Google Cloud Computing.

**V1. Stack tecnológico**

**Programación**:

Python y Jupyter: para los scripts en donde descargamos los archivos csvs que contienen nuestros datos, para posteriormente el análisis de los mismos, crear funciones que automaticen la carga incremental, el ETL y el envío de datos hacia BigQuery en forma de tabla, y para nuestro modelo predictivo de Machine Learning.

Librerías:

Pandas: hemos utilizado Pandas tanto en el EDA, como el ETL y en el modelo de Machine Learning, ya que al trabajar con datos la forma más completa y sencilla de hacerlo es a partir de dataframes.

Scikit-Learn: es una de las bibliotecas mas populares para el aprendizaje automático con Python, y en nuestro caso, hemos utilizado las funcionalidades de regresión lineal y manejo de valores faltante para construir un modelo capaz de predecir la migración neta.

Mathplotlib: esta es una librería utilizada para la creación de graficos, decidimos elegirla porque todos la conocíamos y el uso que le dimos fue durante el EDA.

**Visualización**:

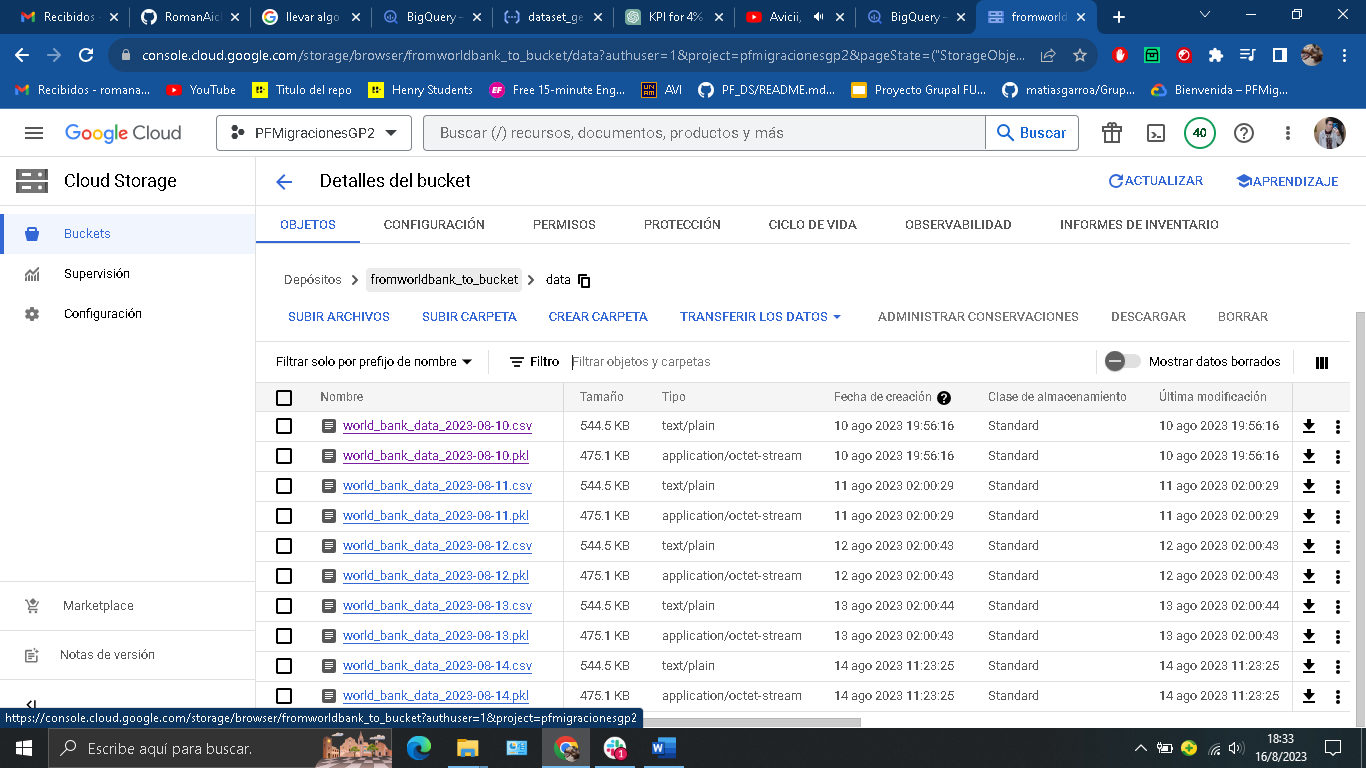
Jupyter: para el EDA.

Power BI: para un análisis más exhaustivo de los datos, la creación de KPIs y el dashboard.

**Cloud** **Computing**:

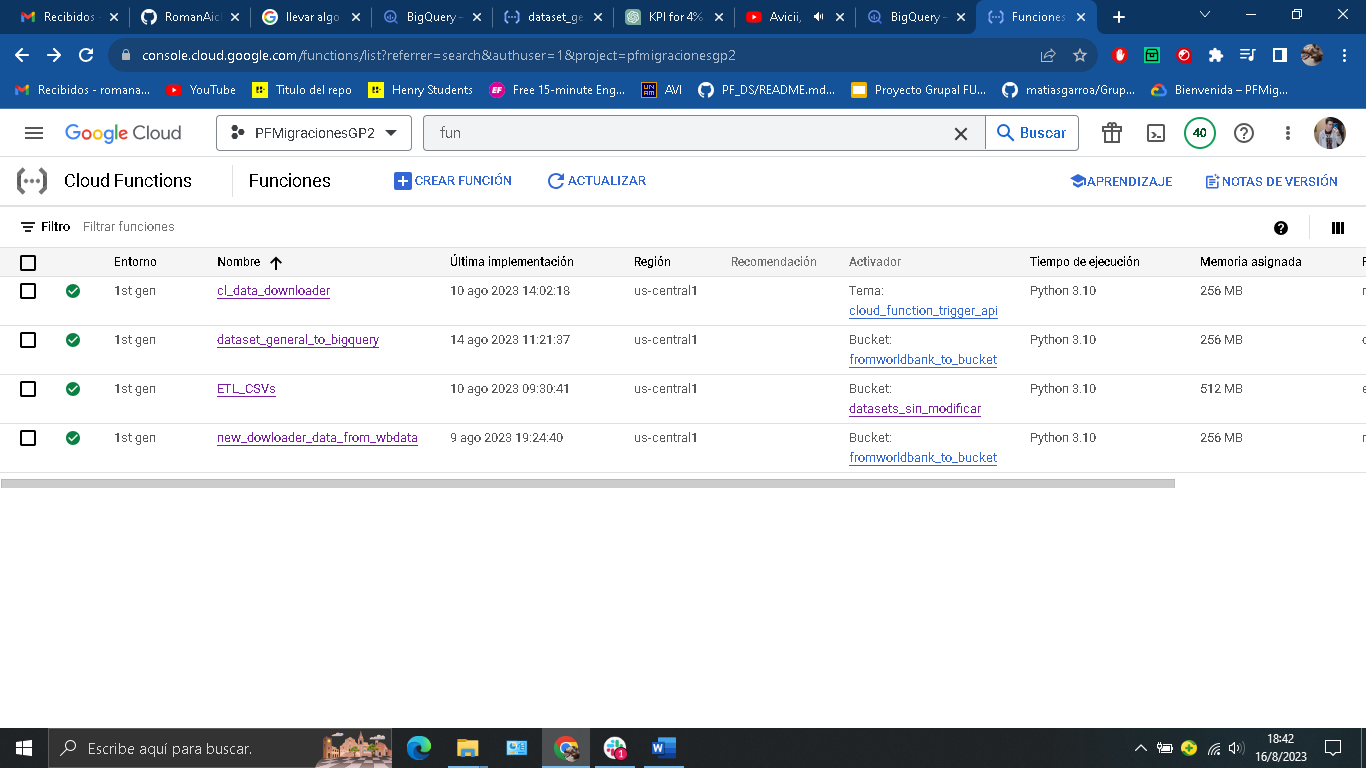
Google Cloud Storage:

Esta tecnología la utilizamos para nuestro Datalake, aquí llegan los datos en bruto para ser almacenados luego de realizarse el ETL.



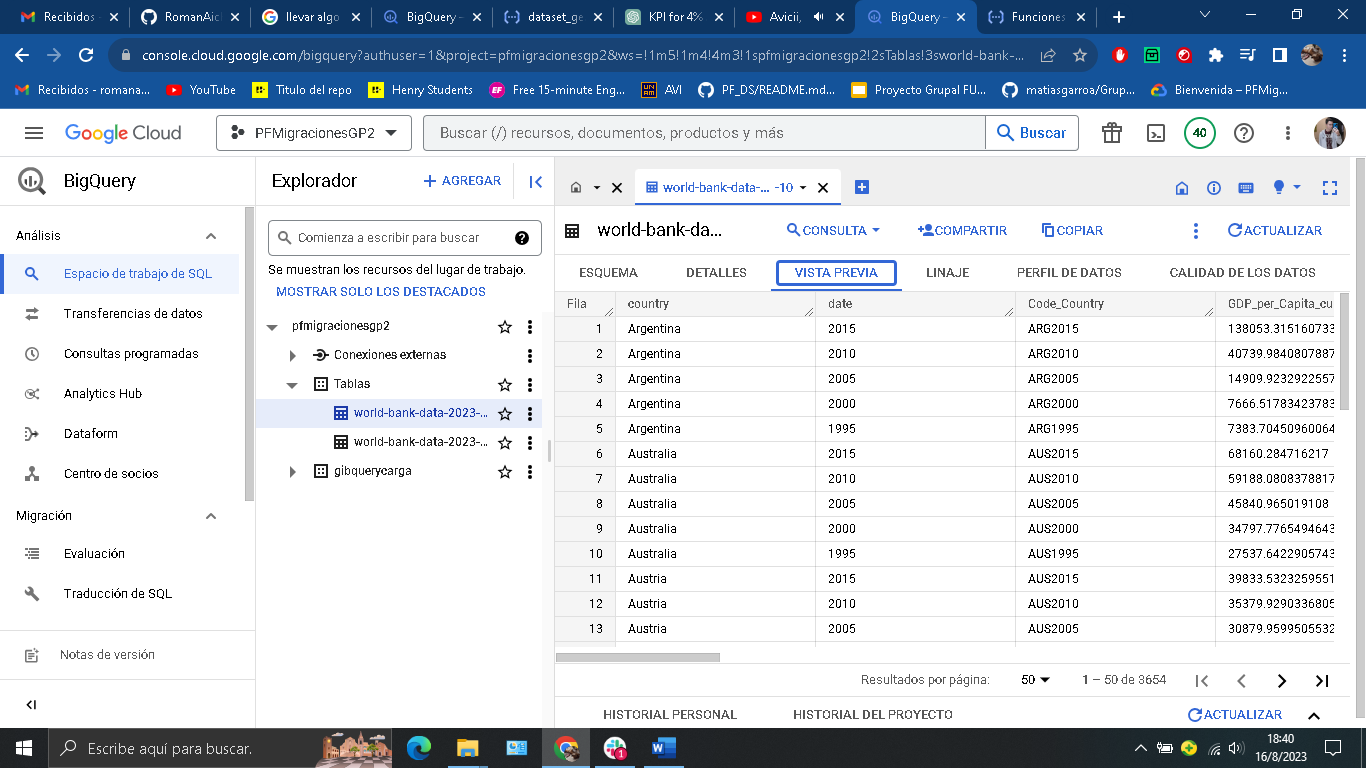
Cloud Functions:

En esta plataforma hemos creado las funciones que automatizan el ciclo completo del dato, en donde cada una hacen el trabajo de un pipeline, empezando por una primera función que realiza la extracción de datos a través de APIs, luego les realiza el ETL y finalmente lo almacena en un bucket de Google Cloud Storage; al terminar eso, inmediatamente de ejecuta la segunda función con la tarea de importar los datos del Datalake, y cargarlos en BigQuery en forma de tabla.



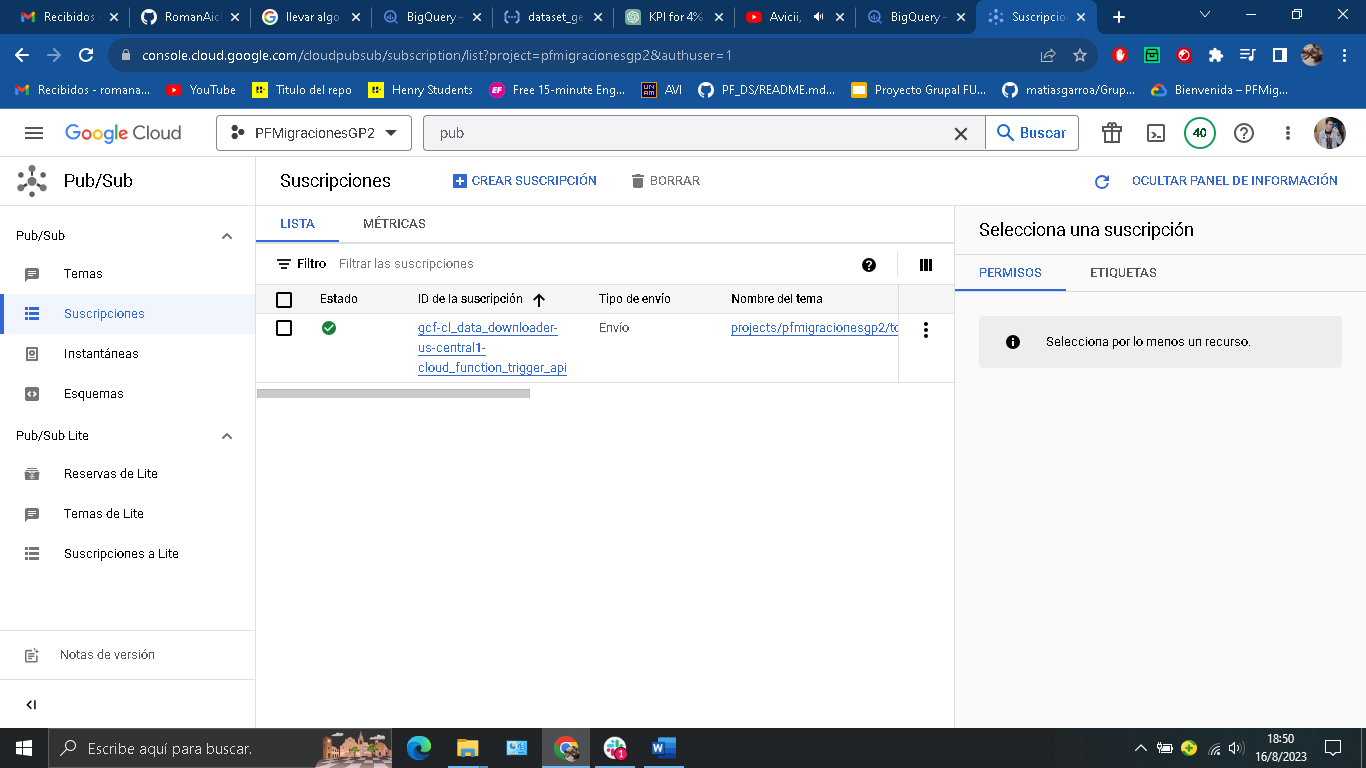
BigQuery:

Aquí, como nombre anteriormente, es donde almacenamos nuestros datos finales, en tablas que mensualmente se actualizarán. BigQuery es lo que vamos a utilizar como Datawherehouse, ya que además de tener una gran simpleza a la hora de aprender a usarlo, también cuenta con SQL como lenguaje de consulta, lo que lo hace sencillo para las personas que estudiamos Ciencia de datos.



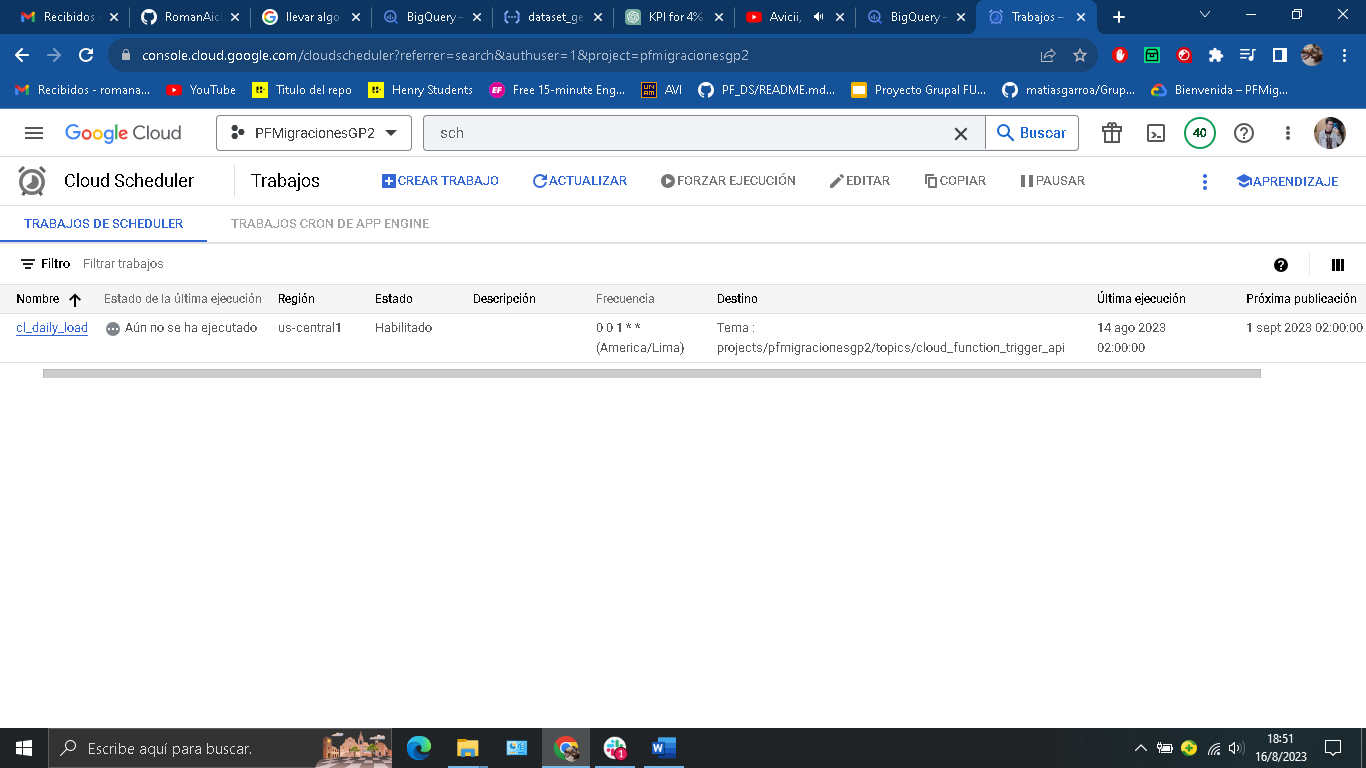
Pub Sub:

Lo utilizamos para establecer una comunicación entre los microservicios.



Cloud Scheduler:

Con este servicio configuramos la activación de la primera función de Cloud Functions, haciendo que se ejecute el primer día de cada mes ya que no sabemos con exactitud cuando habrá actualizaciones de los datos en las páginas, pero si sabemos que suele ser una actualización por año.



**V2. Metodología de trabajo**

Para el trabajo nos dividimos en fases, en donde hacíamos las cosas algunas veces por separados y otras todos juntos.

Fase 1: Indagación en los datos para decidir a qué país o región íbamos a enfocarnos.

Fase 2: Comenzamos a bajar manualmente los datos para ver que teníamos y analizar que podíamos hacer con ellos.

Fase 3: Empezamos a indagar en Google Cloud Platform, para ver que herramientas tenía y decidir cuales íbamos a usar durante el transcurso del trabajo.

Fase 4: Determinamos los datasets a utilizar y comenzamos a realizar el ETL de los mismos en Cloud Functions.

Fase 5: Determinamos métricas e iniciamos los KPIs, aunque más adelante serían modificados.

Fase 6: Automatización de la carga incremental.

Fase 7: Automatización del traspaso de datos al Datawherehouse.

Fase 8: Finalizar los KPIs y volcarlos en un dashboard interactivo.

Fase 9: Creación de un modelo de Machine Learning que pueda predecir las futuras olas migratorias.

Fase 10: Almacenar todo en un repositorio de Github.

**V3. Roles y responsabilidades**

Roman Aichino: Data Engineer

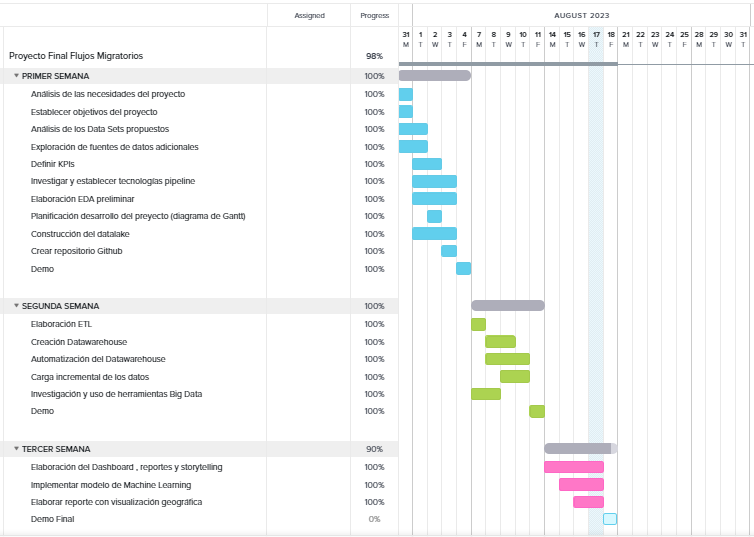
Santiago Babsia: Data Engineer – Machine Learning

Max Jeffer: Data Analytics – Data Engineer

Gabriel Guiffrida: Data Analytics – Machine Learning

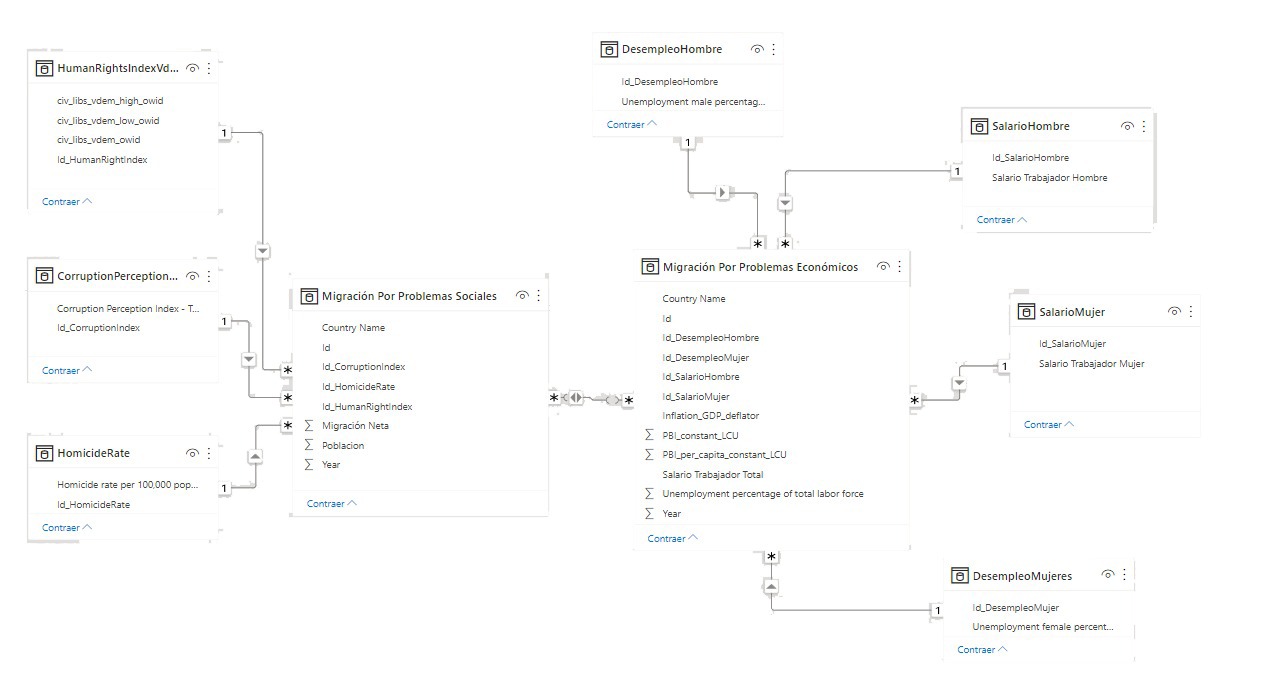
Monserrat Castillo: Data Analytics

**V4. Cronograma de Gantt**



Sprint 2: Diseño de modelo de tablas relacionales, automatización de carga, descarga y actualización de base de datos en ambiente cloud, automatización de normalización de base de datos en ambiente cloud.

I. Diseño de modelo ER (Entidad Relación)



II. Automatización pipeline

Automatización de carga, descarga y actualización de base de datos en ambiente cloud:

Permítanme presentarles la fascinante técnica de carga incremental de datos, un proceso que optimiza el acceso a información actualizada y valiosa. A través de la utilización de diversas APIs, estamos en condiciones de extraer los datos que necesitamos. Este proceso está cuidadosamente programado mediante Google Cloud Function, lo cual nos brinda un flujo continuo de información relevante.

Imagínense, cada mes, este proceso se activa de manera automatizada gracias a Google Schedule. Esto significa que siempre estamos al tanto de los datos más recientes, lo que a su vez nos permite tomar decisiones basadas en información actualizada y precisa.

Una vez que los datos son obtenidos, son dirigidos de forma automática hacia un almacenamiento seguro: nuestro bucket. La gestión de este bucket también está programada a través de una función en Google Cloud, la cual crea las tablas necesarias con los datos preparados y listos para ser utilizados en Google BigQuery.

Este proceso no solo nos brinda datos siempre actualizados, sino que también optimiza nuestra eficiencia al automatizar cada paso del camino, desde la extracción inicial hasta la preparación y almacenamiento en BigQuery. Aprovechamos la tecnología para hacer que la gestión de datos sea más inteligente y eficaz.

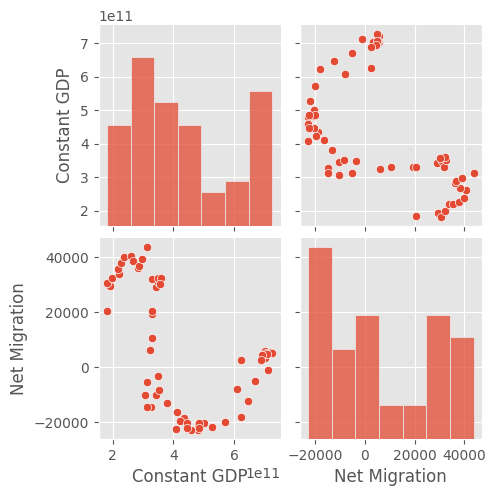
lll. Diccionario de datos

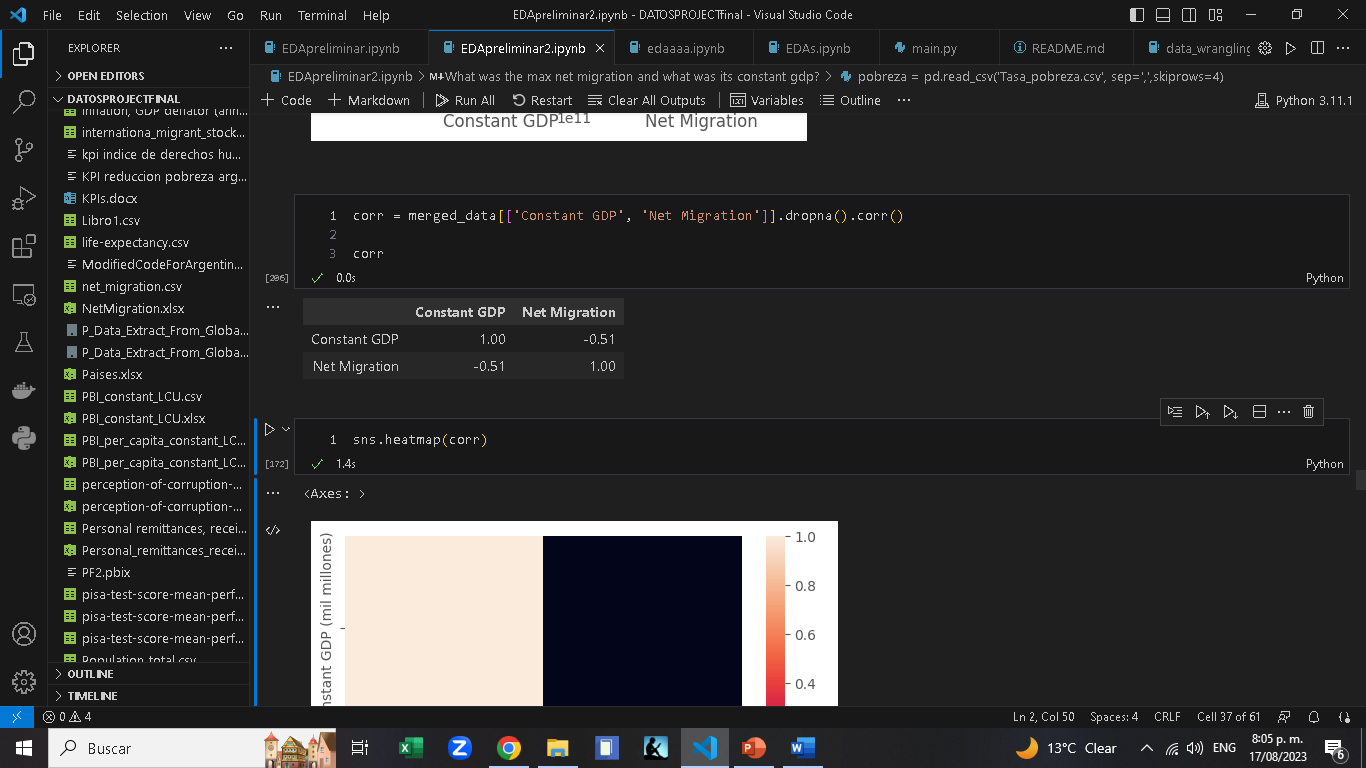
|  |  |  |  |
| --- | --- | --- | --- |
| Nombre de Variable (Columna) | Definición | Código Indicator | Fuente |
| Corruption Perception Index, 2018 | La corrupción es un fenómeno que involucra muchos aspectos diferentes y, por lo tanto, es difícil dar una definición precisa y completa. Sin embargo, en el centro de la mayoría de las definiciones de corrupción está la idea de que un acto corrupto implica el abuso del poder encomendado para beneficio privado. Los ejemplos clásicos incluyen el soborno, el clientelismo y la malversación. Otros ejemplos de corrupción, a menudo más sutiles ya veces incluso legales, incluyen cabildeo y patrocinio. |  | https://ourworldindata.org/corruption |
| GDP (constant LCU) | El PIB es la suma del valor agregado bruto de todos los productores residentes en la economía más los impuestos sobre los productos y menos los subsidios no incluidos en el valor de los productos. Se calcula sin hacer deducciones por depreciación de bienes fabricados o por agotamiento y degradación de los recursos naturales. Los datos están en moneda local constante. | NY.GDP.MKTP.KN | Fuente Datos de cuentas nacionales del Banco Mundial y archivos de datos de cuentas nacionales de la OCDE. |
| GDP (current LCU) | El PIB a precios de comprador es la suma del valor agregado bruto de todos los productores residentes en la economía más los impuestos sobre los productos y menos los subsidios no incluidos en el valor de los productos. Se calcula sin hacer deducciones por depreciación de bienes fabricados o por agotamiento y degradación de los recursos naturales. Los datos están en la moneda local actual. | NY.GDP.MKTP.CN | World Bank national accounts data, and OECD National Accounts data files. |
| GDP per capita (constant LCU) | El PIB per cápita es el producto interno bruto dividido por la población a mitad de año. El PIB a precios de comprador es la suma del valor agregado bruto de todos los productores residentes en la economía más los impuestos sobre los productos y menos los subsidios no incluidos en el valor de los productos. Se calcula sin hacer deducciones por depreciación de bienes fabricados o por agotamiento y degradación de los recursos naturales. Los datos están en moneda local constante. | NY.GDP.PCAP.KN | Fuente Datos de cuentas nacionales del Banco Mundial y archivos de datos de cuentas nacionales de la OCDE. |
| GDP per capita (current LCU) | El PIB per cápita es el producto interno bruto dividido por la población a mitad de año.  Los datos están en la moneda local actual. | NY.GDP.PCAP.CN | Fuente Datos de cuentas nacionales del Banco Mundial y archivos de datos de cuentas nacionales de la OCDE. |
| Homicide Rate, 2021 | Los homicidios, cuando las personas matan intencional e ilegalmente a otros por motivos personales, son los delitos más graves. |  | https://ourworldindata.org/homicides |
| Human Rights Index | Basado en las evaluaciones de expertos y el índice de V-Dem. Capta hasta qué punto las personas están libres de  tortura gubernamental, asesinatos políticos y trabajos forzados; tienen derechos de propiedad; y disfrutar de las libertades de  movimiento, religión, expresión y asociación. La variable va de 0 a 1 (la mayoría de los derechos). |  | https://ourworldindata.org/human-rights |
| Indice de Desarrolo Humano | El Índice de Desarrollo Humano (IDH) es una medida resumida del logro promedio en dimensiones clave del desarrollo humano: una vida larga y saludable, estar bien informado y tener un nivel de vida digno. El IDH es la media geométrica de los índices normalizados para cada una de las tres dimensiones. La dimensión salud se evalúa por la esperanza de vida al nacer, la dimensión educación se mide por medio de los años de escolaridad de los adultos de 25 años y más y los años esperados de escolaridad de los niños en edad de ingresar a la escuela. La dimensión del nivel de vida se mide por el ingreso nacional bruto per cápita. El IDH utiliza el logaritmo de los ingresos para reflejar la disminución de la importancia de los ingresos a medida que aumenta el INB. Las puntuaciones de los tres índices de las dimensiones del IDH se agregan luego en un índice compuesto utilizando la media geométrica. Consulte las Notas técnicas para obtener más detalles. La variable va de 0 a 1 |  | https://hdr.undp.org/data-center/human-development-index#/indicies/HDI |
| Inflation, Consumer Prices (Annual %) | La inflación medida por el índice de precios al consumidor refleja el cambio porcentual anual en el costo para el consumidor promedio de adquirir una canasta de bienes y servicios que puede ser fijo o cambiar en intervalos específicos, como anualmente. Generalmente se utiliza la fórmula de Laspeyres. | FP.CPI.TOTL.ZG | Fuente Fondo Monetario Internacional, Estadísticas Financieras Internacionales y archivos de datos. |
| Inflation, GDP Deflator (Annual %) | La inflación, medida por la tasa de crecimiento anual del deflactor implícito del PIB, muestra la tasa de cambio de precios en la economía en su conjunto. El deflactor implícito del PIB es la relación entre el PIB en moneda local corriente y el PIB en moneda local constante. | NY.GDP.DEFL.KD.ZG | Fuente Datos de cuentas nacionales del Banco Mundial y archivos de datos de cuentas nacionales de la OCDE. |
| International migrant stock (% of population) | El stock de migrantes internacionales es el número de personas nacidas en un país diferente a aquel en el que viven. También incluye a los refugiados. Los datos utilizados para estimar el stock de migrantes internacionales en un momento determinado se obtienen principalmente de los censos de población. Las estimaciones se derivan de los datos sobre la población nacida en el extranjero: personas que tienen residencia en un país, pero nacieron en otro. Cuando no se dispone de datos sobre la población nacida en el extranjero, se utilizan como estimaciones los datos sobre la población extranjera, es decir, las personas que son ciudadanos de un país distinto del país en el que residen. Después de la desintegración de la Unión Soviética en 1991, las personas que vivían en uno de los países recientemente independizados que nacieron en otro fueron clasificadas como inmigrantes internacionales. Las estimaciones de población migrante en los nuevos estados independientes a partir de 1990 se basan en el censo de 1989 de la Unión Soviética. Para los países con información sobre el stock de migrantes internacionales para al menos dos puntos en el tiempo, se utilizó la interpolación o extrapolación para estimar el stock de migrantes internacionales el 1 de julio de los años de referencia. Para los países con una sola observación, las estimaciones para los años de referencia se derivaron utilizando las tasas de cambio en la población migrante en los años anteriores o posteriores a la única observación disponible. Se utilizó un modelo para estimar los migrantes de países que no tenían datos. | SM.POP.TOTL.ZS | United Nations Population Division, Trends in Total Migrant Stock: 2008 Revision. |
| Net migration | La migración neta es el número de inmigrantes menos el número de emigrantes, incluidos ciudadanos y no ciudadanos, para el período de cinco años. | SM.POP.NETM | United Nations Population Division, World Population Prospects 2008. |
| Personal remittances, received (% of GDP) | Las remesas personales comprenden las transferencias personales y la remuneración de los empleados. Las transferencias personales consisten en todas las transferencias corrientes en efectivo o en especie realizadas o recibidas por hogares residentes hacia o desde hogares no residentes. Las transferencias personales incluyen, por lo tanto, todas las transferencias corrientes entre personas físicas residentes y no residentes. La remuneración de los empleados se refiere a los ingresos de los trabajadores fronterizos, de temporada y otros trabajadores a corto plazo que están empleados en una economía en la que no son residentes y de los residentes empleados por entidades no residentes. Los datos son la suma de dos elementos definidos en la sexta edición del Manual de Balanza de Pagos del FMI: transferencias personales y remuneración de los empleados. | BX.TRF.PWKR.DT.GD.ZS | Fuente Estimaciones del personal del Banco Mundial basadas en datos de la balanza de pagos del FMI y estimaciones del PIB del Banco Mundial y la OCDE. |
| Population, total | La población total se basa en la definición de población de facto, que cuenta a todos los residentes independientemente de su estatus legal o ciudadanía. Los valores que se muestran son estimaciones de mitad de año. | SP.POP.TOTL | (1) United Nations Population Division. World Population Prospects: 2022 Revision. (2) Census reports and other statistical publications from national statistical offices, (3) Eurostat: Demographic Statistics, (4) United Nations Statistical Division. Population and Vital Statistics Report (various years), (5) U.S. Census Bureau: International Database, and (6) Secretariat of the Pacific Community: Statistics and Demography Programme. |
| Poverty headcount ratio at national poverty lines (% of population) | La tasa nacional de recuento de la pobreza es el porcentaje de la población que vive por debajo de la(s) línea(s) nacional(es) de pobreza. Las estimaciones nacionales se basan en estimaciones de subgrupos ponderados por población a partir de encuestas de hogares. Para las economías cuyos datos provienen de EU-SILC, el año informado es el año de referencia de ingresos, que es el año anterior al año de la encuesta. | SI.POV.NAHC | World Bank, Poverty and Inequality Platform. Data are compiled from official government sources or are computed by World Bank staff using national (i.e. country–specific) poverty lines. |
| Unemployment, female (% of female labor force) (modeled ILO estimate) | El desempleo se refiere a la proporción de la fuerza laboral que no tiene trabajo pero que está disponible y buscando empleo. | SL.UEM.TOTL.FE.ZS | International Labour Organization, ILOSTAT database. Data retrieved in September 2018. |
| Unemployment, male (% of male labor force) (national estimate) | El desempleo se refiere a la proporción de la fuerza laboral que no tiene trabajo pero que está disponible y buscando empleo. Las definiciones de fuerza laboral y desempleo difieren según el país. | SL.UEM.TOTL.MA.NE.ZS | International Labour Organization. “Labour Force Statistics database (LFS)” ILOSTAT. Accessed April 25, 2023. https://ilostat.ilo.org/data/. |
| Unemployment, total (% of total labor force) (modeled ILO estimate) | El desempleo se refiere a la proporción de la fuerza laboral que no tiene trabajo pero que está disponible y buscando empleo. | SL.UEM.TOTL.ZS | International Labour Organization. “ILO Modelled Estimates and Projections database (ILOEST)” ILOSTAT. Accessed April 25, 2023. https://ilostat.ilo.org/data/. |
| Wage and salaried workers, female (% of female employment) (modeled ILO estimate) | Los trabajadores asalariados (empleados) son aquellos trabajadores que ocupan el tipo de trabajos definidos como "trabajos de empleo remunerado", donde los titulares tienen contratos de trabajo explícitos (escritos u orales) o implícitos que les otorgan una remuneración básica que no depende directamente de los ingresos de la unidad para la que trabajan. | SL.EMP.WORK.FE.ZS | International Labour Organization, ILOSTAT database. Data retrieved in September 2018. |
| Wage and salaried workers, male (% of male employment) (modeled ILO estimate) | Los trabajadores asalariados (empleados) son aquellos trabajadores que ocupan el tipo de trabajos definidos como "trabajos de empleo remunerado", donde los titulares tienen contratos de trabajo explícitos (escritos u orales) o implícitos que les otorgan una remuneración básica que no depende directamente de los ingresos de la unidad para la que trabajan. | SL.EMP.WORK.MA.ZS | International Labour Organization. “ILO modelled estimates database” ILOSTAT. Accessed January 2021. https://ilostat.ilo.org/data/. |
| Wage and salaried workers, total (% of total employment) (modeled ILO estimate) | Los trabajadores asalariados (empleados) son aquellos trabajadores que ocupan el tipo de trabajos definidos como "trabajos de empleo remunerado", donde los titulares tienen contratos de trabajo explícitos (escritos u orales) o implícitos que les otorgan una remuneración básica que no depende directamente de los ingresos de la unidad para la que trabajan. | SL.EMP.WORK.ZS | International Labour Organization. “ILO modelled estimates database” ILOSTAT. Accessed January 2021. https://ilostat.ilo.org/data/. |

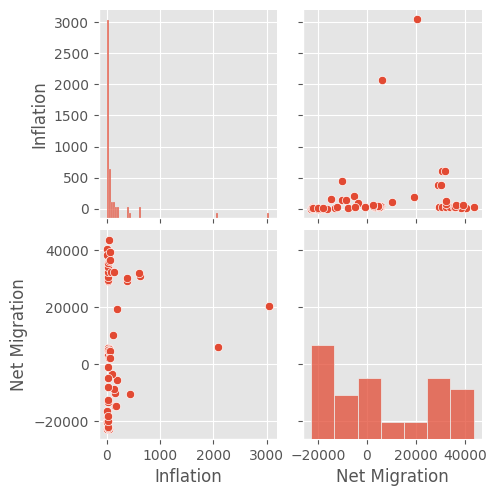
Sprint 3: Análisis de datos, machine learning y visualización de datos para diseño de reportes, dashboard, y verificar funcionalidad.

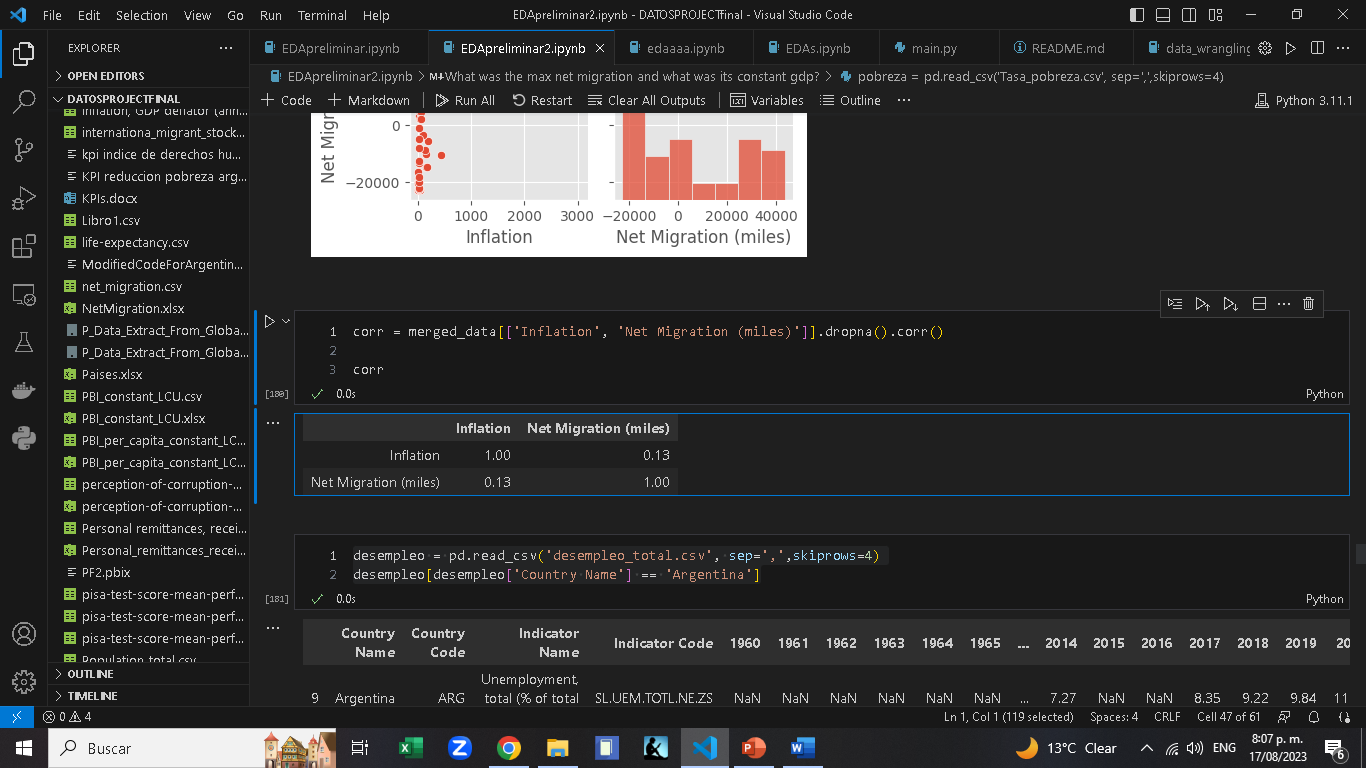
I. Diseño de reportes

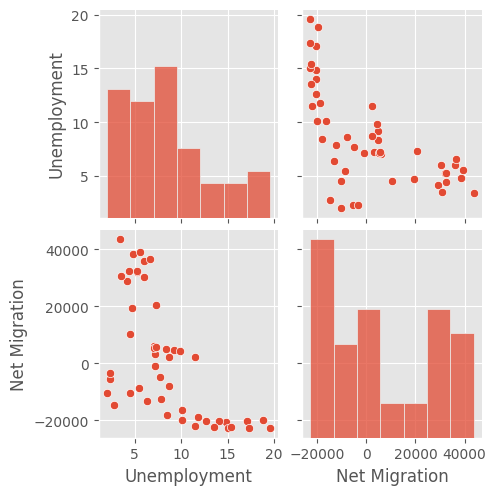
**Pairplot Migración Neta Vs PBI en Términos Constantes**

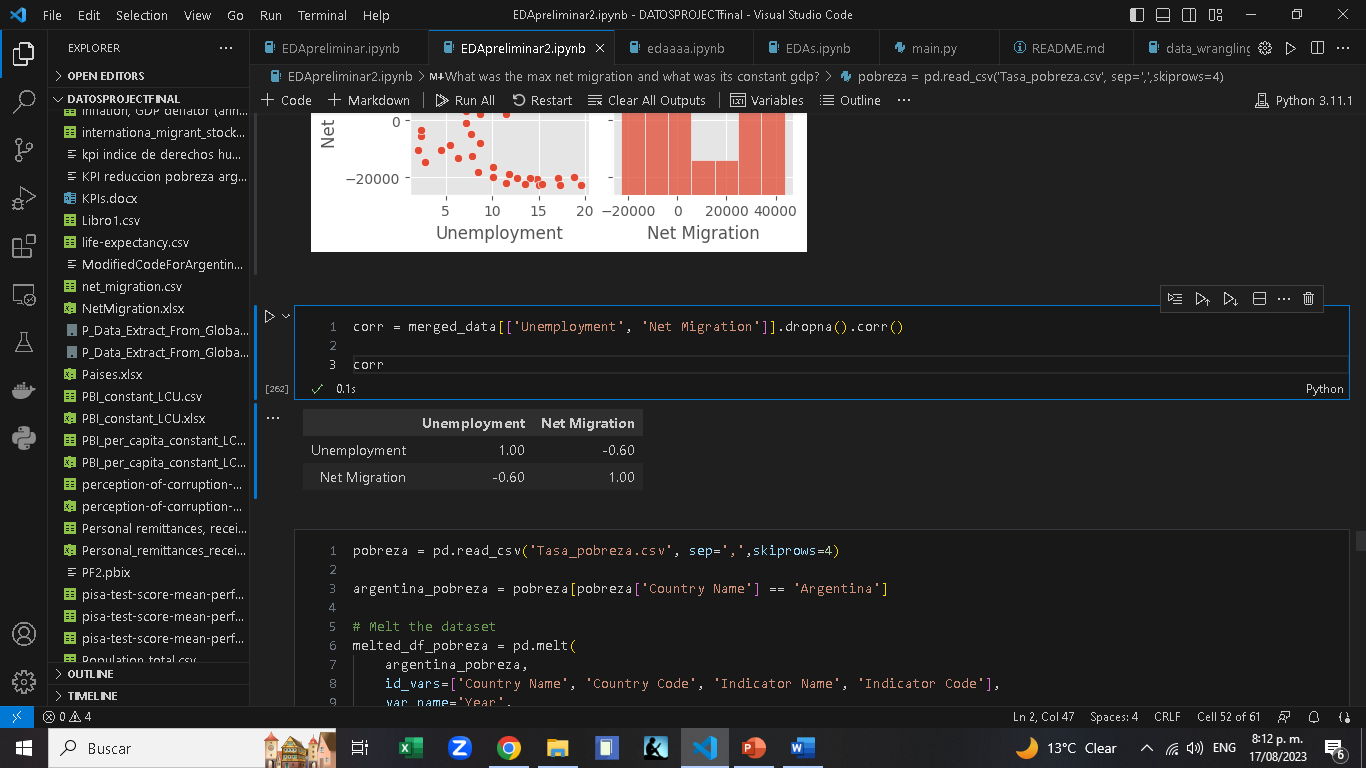












I.1 métricas de monitoreo

Índice de calidad de vida = (Índice de percepción de corrupción + Índice de desarrollo humano) / 2

Índice de estabilidad económica = (PBI per cápita (nominal) / Inflación) \*100

Tasa de reducción de pobreza = ((Tasa de pobreza inicial – tasa de pobreza actual) / Tasa de pobreza inicial) \*100

Índice de seguridad y migración = ()

Tasa de emigración = (Migración neta / Población total) \*100

Tasa de fuga de cerebros = (Profesionales calificados emigrados / Total de profesionales calificados) \*100

Índice de derechos humanos y gobernanza = (Índice de percepción de corrupción / Índice de desarrollo humano)

I.2 KPIs

KPI Índice de calidad de vida: aumentar ese índice en un 3% anual.

KPI Índice de derechos humanos y gobernanza: aumentar el índice en un 4% anual.

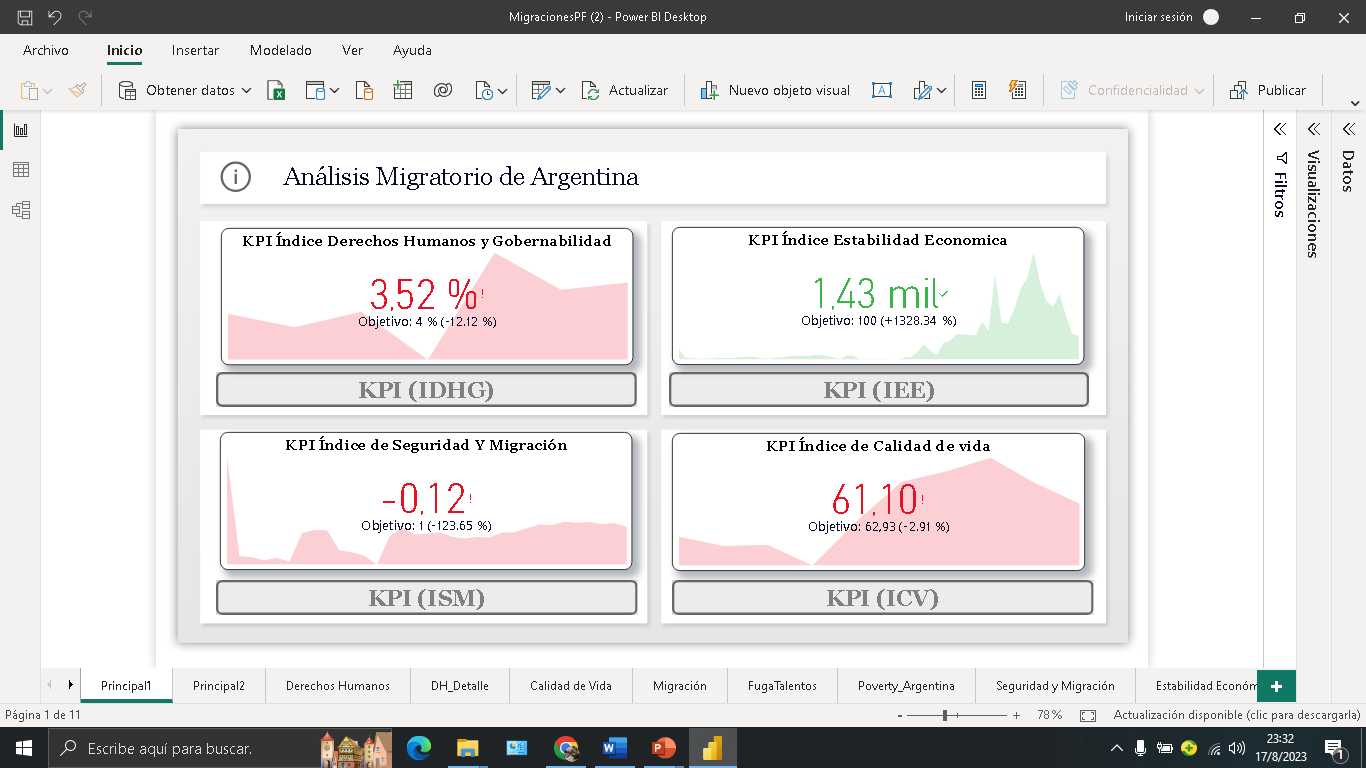
KPI Tasa de reducción de pobreza: reducirla hasta llegar al 27% en 2021.

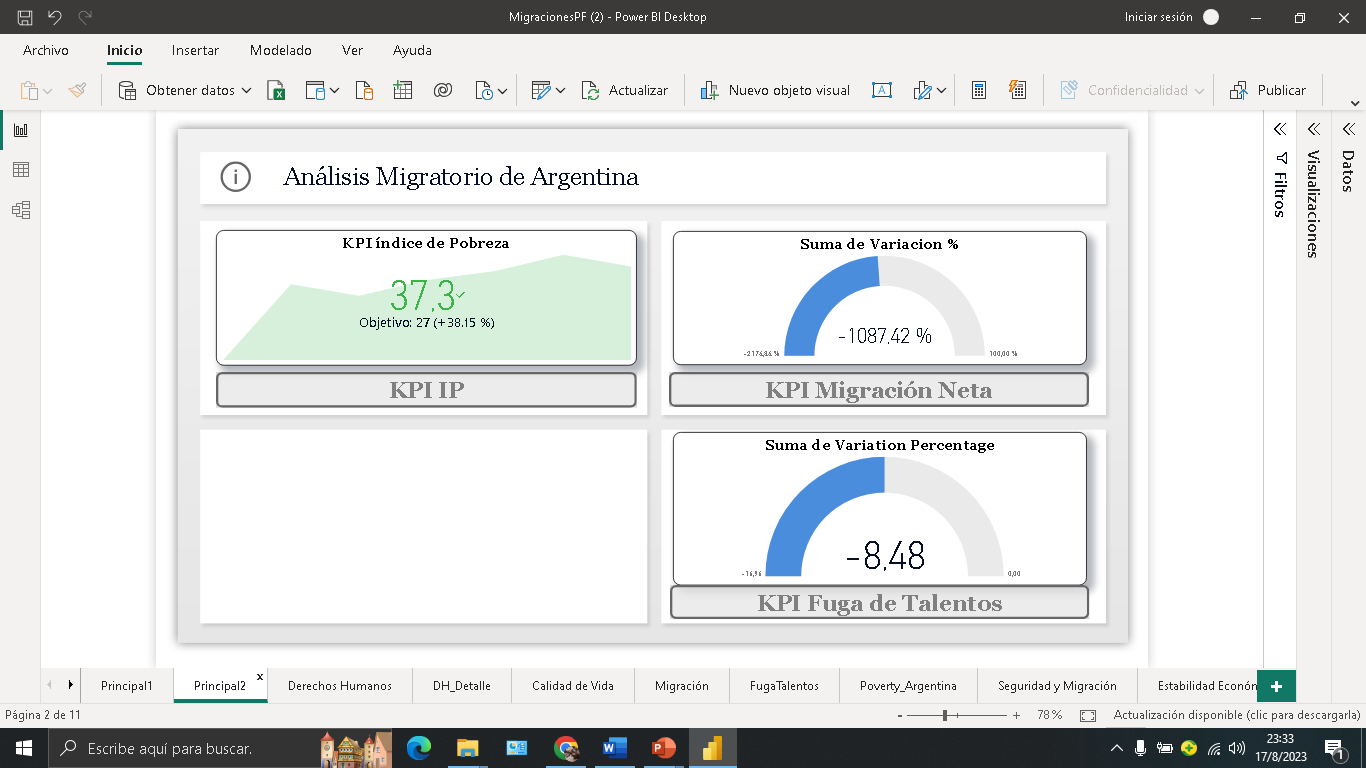
KPI Índice de seguridad y migración: aumentarla en un 5% por año.

KPI Índice de estabilidad económica: mejorarlo en un 5%

II. Dashboards

Estas son las dos paginas principales, con los links a las explicaciones de cada uno de los KPIs





lll. Implementación modelos machine Learning

En este modelo estamos interesados en comprender cómo los cambios en la tasa de desempleo pueden afectar la migración neta en Argentina. Para abordar esta pregunta, se ha utilizado la regresión lineal, una metodología que modela la relación entre dos variables, en este caso, la tasa de desempleo y la migración neta, mediante una línea recta.

El Proceso:

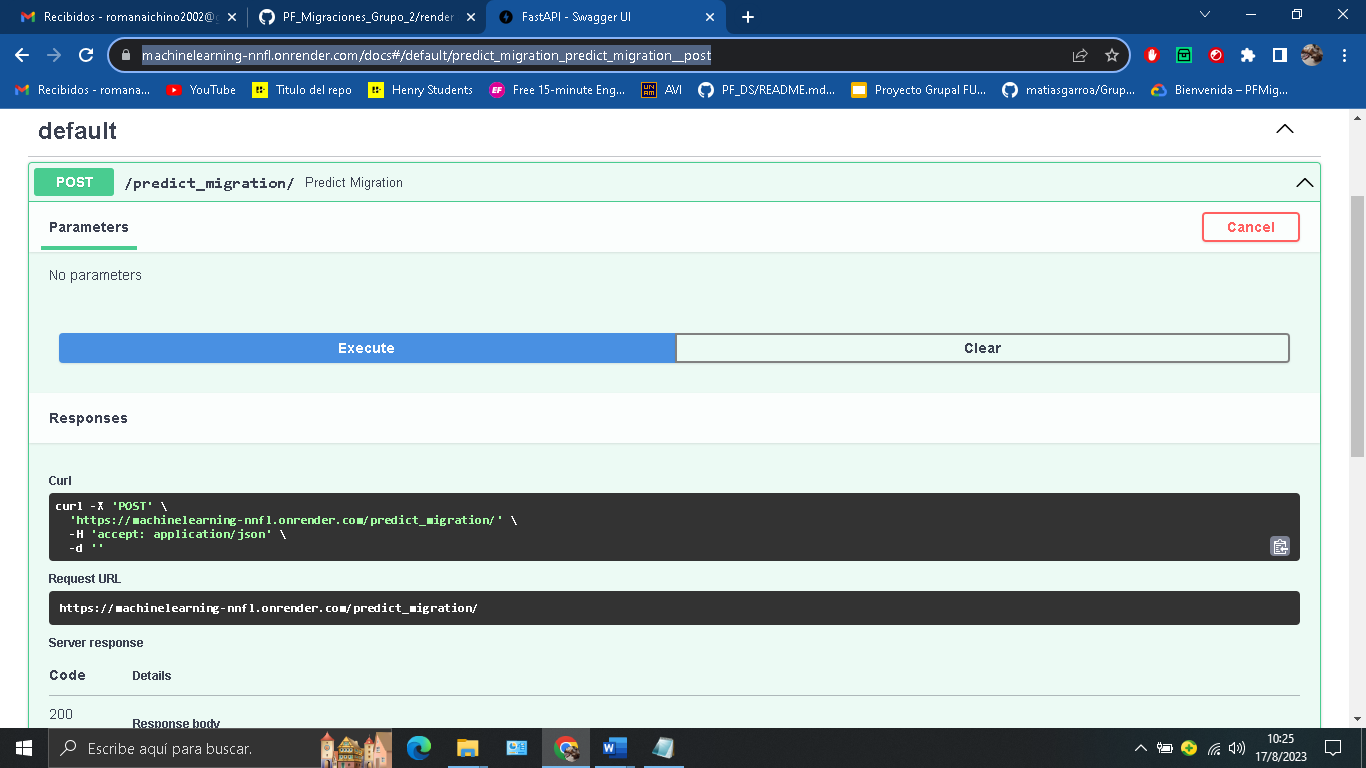
Carga y Preprocesamiento de Datos: Utilizamos Pandas para cargar los datos desde el archivo 'datos.csv'. Luego, seleccionamos solo los datos relacionados con Argentina. La tasa de desempleo es nuestra variable independiente, y utilizamos SimpleImputer para manejar los valores faltantes mediante la imputación de la media.

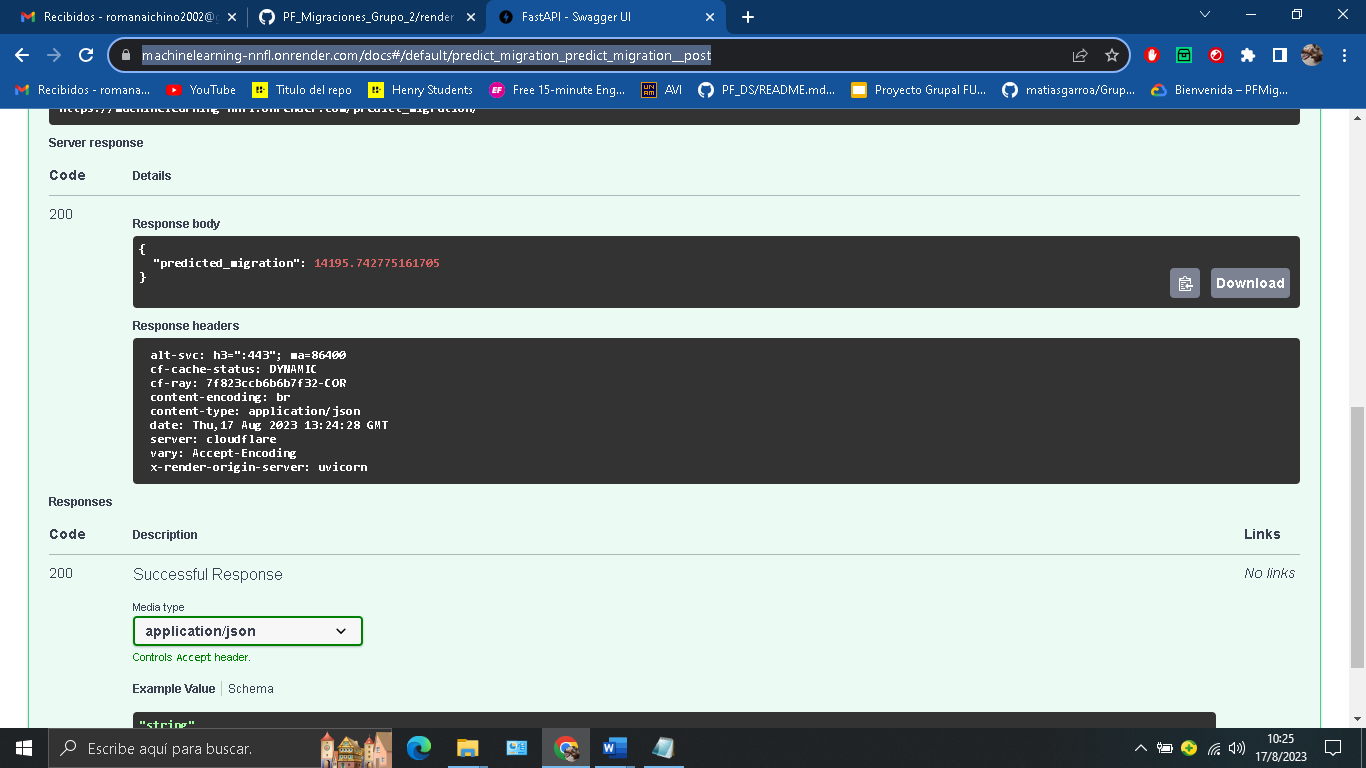
Creación del Modelo de Regresión Lineal: Hemos aprovechado la funcionalidad de regresión lineal de scikit-learn. Este modelo matemático se adapta a los datos de Argentina, capturando la relación entre la tasa de desempleo y la migración neta.

Predicción y Resultados: Supongamos que queremos predecir la migración neta en 2023. Introducimos una tasa de desempleo proyectada y dejamos que el modelo haga el trabajo. La migración neta esperada para 2023 se almacena en ‘predicted\_migration\_2023’.

El modelo esta subido a render:

Link: <https://machinelearning-nnfl.onrender.com/docs#/default/predict_migration_predict_migration__post>





CONCLUSION

Con base en la información proporcionada, se puede concluir que el proyecto liderado por GMMRS CONSULTING en colaboración con el Gobierno de la República Argentina ha logrado implementar un enfoque integral y tecnológicamente avanzado para abordar los desafíos y oportunidades que presentan los flujos migratorios en el país. El proyecto se ha enfocado en la recolección, procesamiento y análisis de datos relacionados con la migración, con el objetivo de proporcionar una comprensión profunda de este fenómeno y brindar herramientas para la toma de decisiones informadas.

La estrategia adoptada ha sido la creación de un pipeline robusto y automatizado que abarca desde la carga inicial de datos hasta la generación de predicciones y visualizaciones accesibles a través de una plataforma en línea. Este enfoque permite a las partes interesadas, tanto dentro del gobierno como en el público en general, acceder a información actualizada y relevante sobre los flujos migratorios en Argentina.

En términos de habilidades, el proyecto ha fortalecido tanto las habilidades técnicas como las habilidades blandas del equipo de GMMRS CONSULTING. El uso de tecnologías de pipeline, la automatización de procesos y la implementación de modelos de machine learning han consolidado las habilidades técnicas del equipo en el campo de la Ciencia de Datos. Además, la presión para cumplir con los objetivos del proyecto ha mejorado la comunicación bajo circunstancias demandantes.

En conclusión, el proyecto ha sido exitoso en su objetivo de proporcionar una herramienta integral y de vanguardia para comprender, analizar y predecir los flujos migratorios en Argentina. La colaboración entre GMMRS CONSULTING y el Gobierno de la República Argentina ha resultado en un valioso recurso que beneficia tanto a las autoridades gubernamentales como a la sociedad en general, al permitir tomar decisiones fundamentadas en relación con la migración y sus implicaciones. El despliegue del Modelo de Predicción y el acceso al dashboard demuestran un compromiso sólido con la aplicación práctica de la Ciencia de Datos para abordar problemas sociales complejos.