



ONG - **Soy Henry**

Reporte Semana #1

Propuesta de proyecto

Indice

Página 3. Entendimiento de la situación actual

Página 4. Objetivos: específicos del trabajo y del grupo

Página 5. Alcance y fuera de alcance

Página 5. Repositorio

Página 6. Solución propuesta

- Stack tecnológico
- Metodología de trabajo
- Roles y responsabilidades
- Cronograma general y estimación de esfuerzos

Página 8. KPI's a utilizar

Entendimiento de la situación actual

El informe requerido por nuestro cliente es sobre la temática de los Flujos Migratorios (FM). Los FM hacen referencia al número de migrantes que llegan a una zona o parten de una zona en el transcurso de un período específico de tiempo. Dichas zonas pueden ser países, regiones, ciudades o continentes. El término surgió en reemplazo de otros términos con una connotación más despectiva, tales como “masa” u “oleada” de inmigrantes.

Los fenómenos de la migración y la movilidad humana son de característica innata al ser humano, de índole milenaria y han afectado casi sin excepción a todas las sociedades del mundo. La migración es un tema complejo, y como tal puede ser distorsionado en grados alarmantes por la información errónea y la politización.

El entendimiento de los FM desde la perspectiva que nos provee la Ciencia de Datos será el criterio característico de nuestro enfoque, sin caer en sesgos ideológicos ni en opiniones de índole personal.

Creemos que explorar y exponer de forma clara, precisa y científica los FM tiene una utilidad enorme para la sociedad en general, las empresas, los policy makers y todos los agentes que son parte de este mundo, que de forma directa o indirecta se ven afectados por el fenómeno.

En nuestra investigación preliminar hemos podido determinar que las características de los FM varían en función de los lugares, las épocas, las políticas de los países, aspectos demográficos, sociopolíticos, ambientales y económicos, entre otros. Exploraremos en detalles todos estos campos incrementando nuestro entendimiento y reservándonos la prerrogativa de modificar el enfoque del informe a medida que nuestro conocimiento y expertise sobre el fenómeno de estudio se profundice e incrementa.

Objetivos: del trabajo y del grupo

En una primera etapa exploraremos toda la data y trataremos de explicar cuáles son los features determinantes que afectan sobre los FM, los comprenderemos en profundidad y determinaremos cuáles son los más importantes. Haremos análisis de flujos por zonas y continentes, para comprender a grandes rasgos la escala, dirección y frecuencia de los movimientos.

Posteriormente tamizaremos dichos datos a nivel países, y haremos un exhaustivo estudio de los rasgos característicos de ellos, clasificándolos en diversas categorías según su cantidad neta de migración.

Posteriormente realizaremos profundas tareas de analítica y aplicaremos Modelos de Machine Learning (ML) tratando de responder, entre otras, a las siguientes preguntas:

- *¿Qué hace que un emigrante se vaya de un país?*
- *¿Cuáles son las características predominantes de los países más elegidos por los inmigrantes?*
- *¿Qué consecuencias generan FM en ambos tipos de países?*
- *¿Cómo podemos medir ese efecto?*
- *¿Cuáles son los rasgos característicos de los países con FM netos positivos y negativos?*
- *¿Qué hace que un país tenga FM netos cercanos a cero?*

Producto de todo este trabajo, generaremos un informe detallado de los descubrimientos que nuestro equipo de analistas realizará, con las respectivas conclusiones e insights que nos provean los datos.

Crearemos una base de datos única en el mundo que contendrá toda esta información y se actualizará automáticamente.

Además crearemos un interfaz tecnológica que permita acceder de forma remota no sólo a la base de datos, sino que también a métricas, estadísticas e indicadores puntuales que desarrollaremos según nuestros descubrimientos futuros.

Alcance y fuera de alcance

Inicialmente el alcance del proyecto se extiende, llegado el deadline pactado el 28/2, a entregar al cliente los siguientes productos:

- Acceso al repositorio privado con toda la información del proceso
- Informe detallado de la temática
- +5 KPIs que consideremos determinantes, con su explicación e información de uso e interpretación
- Dashboard Interactivo con métricas, estadísticas y KPIs
- Modelo de ML learning entrenado, con acceso al código e instructivo paso a paso del proceso
- Reunión de Reporte con todo el equipo para presentar los productos entregables, una presentación step by step de todo el proceso y la exposición de nuestros analistas

Este proyecto no contempla:

- La previsión rigurosa de los flujos futuros de FM
- Modificaciones en los requerimientos realizados más allá de la segunda semana de trabajo
- Dar información a los refugiados, solicitantes de asilo y desplazados por conflictos políticos y bélicos respecto de cursos de acción ante su situación

Repositorio

El repositorio oficial de Analytic Hound para este proyecto ya se encuentra creado y será entregado al cliente una vez que toda la documentación se encuentre debidamente pusheada y lista para ser consumida.

Solución propuesta

Stack tecnológico

Utilizaremos una arquitectura que contemple las herramientas que se ven en el diagrama siguiente. Nos reservamos la facultad de modificar en el futuro el stack, en caso de ser necesario para cuestiones operativas imprevistas.

PYTHON
PANDAS
AMAZON

DETA CLOUD
FASTAPI
POWERBI

STREAMLIT
VISUAL STUDIO
CODE

Metodología de trabajo

La metodología de trabajo está apoyada en el marco de trabajo para desarrollo de software SCRUM. Nuestros miembros están capacitados y entrenados en la metodología y utilizaremos varias de sus herramientas para coordinar efectivamente nuestras acciones. Principalmente nos centraremos en:

Sprint Planning Meeting: utilizaremos esta herramienta al inicio de cada Sprint semanal con la totalidad del equipo. Inspeccionaremos el Product Backlog (las tareas, requerimientos y las funcionalidades que el proyecto requiere ese semana) y seleccionaremos los items específicos en los que trabajaremos durante el Sprint semanal.

Daily Scrum: diariamente, por la mañana de forma previa a la reunión con nuestro asesor externo Gonzalo Posse, el equipo se reunirá en este espacio. Durante no más de 15 minutos se hará un seguimiento del proyecto en el que se controlará el cumplimiento de las tareas asumidas. Además se pactarán los objetivos de trabajo diario y se analizan los posibles problemas que hayan limitado o impedido directamente el cumplimiento de los objetivos.

Sprint Review: al final del Sprint semanal, luego de la reunión con nuestro PO Jonathan Deiloff, el equipo se reunirá para evaluar la performance en forma retrospectiva. El objetivo es reflexionar sobre el rendimiento del Sprint e identificar posibilidades de mejora para el próximo Sprint Semanal.

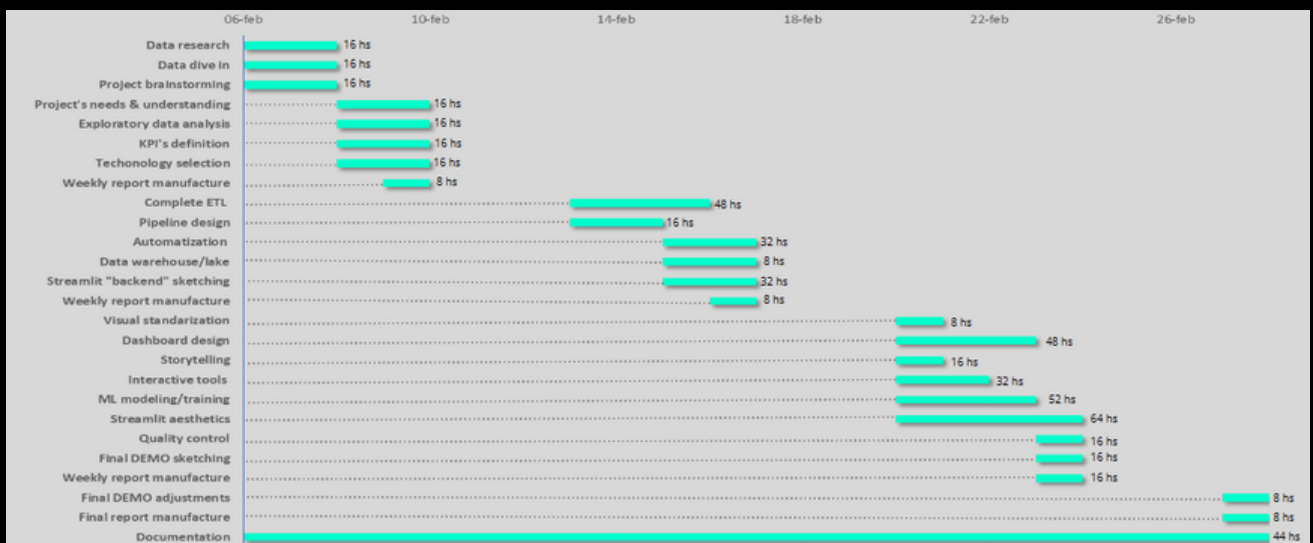
Solución propuesta

Roles y responsabilidades

Nuestro equipo cuenta con diversos perfiles profesionales y de toda la gama en la industria de Data.

- Functional Analyst - Mauro Pini
- Data Engineer - Lucas Rodriguez
- Data Engineer - Eugenia Ball
- Data Analyst - Belén Zapata
- Data Scientist - Alan Mysler

Cronograma general y estimación de esfuerzos



KPI's a utilizar

Según el informe interno preliminar del equipo de Analytics, las hipótesis iniciales de trabajo para la formulación de los KPIs arrojan varios resultados. A medida que el workflow avance, podremos agregar KPIs no considerados inicialmente y/o eliminar alguno de los propuestos en esta etapa inicial del desarrollo del proyecto.

La metodología de determinación de los KPI sigue el concepto de los reconocidos objetivos SMART. El acrónimo refiere a "Specific" (específico), "Measurable" (mensurable), "Attainable" (alcanzable), "Relevant" (relevante) y "Time-bound" (con límite de tiempo).

NOTA: para los KPIs medidos en porcentaje, preferimos la notación decimal en lugar de inscribir el porcentaje en base 100. Ej: 0.2 = 20%. Esta decisión facilitará la manipulación posterior de la información.

- KPI_1 = Aumento de la migración neta anual, medido en cantidad (diferencia entre cantidad de personas que inmigran a un país y cantidad de personas que emigran de ese país entre dos períodos).

$$KPI_1 = MN(n) - MN(n - 1)$$

$MN(n)$ - > Migración neta para el período (n)

$MN(n-1)$ - > Migración neta para el período (n)

Ej: KPI_1 = 150000 implica que el objetivo a alcanzar es que haya un FM neto de 150000 entrantes al país por año.

KPI's a utilizar

- KPI_2 = Aumento de la migración neta anual, medido en porcentaje (variación porcentual respecto del MN respecto al período anterior).

$$KPI_2 = \frac{MN(n) - MN(n - 1)}{MN(n - 1)}$$

Ej: KPI_2 = 0.12 implica que el objetivo a alcanzar es que haya un incremento porcentual del 12% anual del FM neto.

- KPI_3 = Aumento porcentual anual de inmigración calificada. Este KPI mide la variación porcentual anual de los inmigrantes que un país recibe, que tienen grados de escolaridad mayores al secundario.

$$KPI_3 = \frac{MC(n) - MC(n - 1)}{MC(n - 1)}$$

Ej: KPI_3 = 0.15 implica que el objetivo a alcanzar es que haya un incremento porcentual del 15% de la cantidad de inmigrantes calificados.

KPI's a utilizar

- KPI_4 = Incremento del índice de condiciones deseadas por los migrantes en un X%. Este índice está sujeto a especificación posterior, ya que es una basket de los features que encontremos más relevantes a la hora de explicar la recepción de FM en los países que más MN reciban.

$$KPI_4 = \sum_{i=1}^m \left(\frac{Fi(n) - Fi(n-1)}{Fi(n-1)} \right) / m$$

Este KPI nos provee el promedio ponderado de las variaciones porcentuales entre el período n y el período n-1, para los "m" indicadores elegidos. Provee una estimación eficaz del grado de deseabilidad del país para recibir FM.

Ej: KPI_4 = 0.12 implica que el objetivo a alcanzar es que haya un incremento porcentual del 12% anual del indicador. Esto se daría, por ejemplo, en un escenario con 3 Features seleccionados donde sus valores de variación porcentual anual sean: F1=0.05, F2=0.15 y F3=0.16.

-
- KPI_5 = Disminución del índice de condiciones no deseadas por los migrantes en un X%. Este índice está sujeto a especificación posterior ya que es una basket de los features que encontremos más relevantes para explicar los bajos niveles de recepción de FM. La lógica es similar a la de KPI_4, solo que esta canasta mide los features que más explican un FM negativo en un país.

$$KPI_5 = \sum_{i=1}^m \left(\frac{Fi(n) - Fi(n-1)}{Fi(n-1)} \right) / m$$

En este caso un KPI decreciente en el tiempo es deseable porque indica que las condiciones adversas, son menores. Lo óptimo sería que KPI_5 tomará signos negativos. Intuitivamente se puede pensar que una desaceleración en la tasa de inflación, una disminución del desempleo y de los índices de inseguridad son deseables, cuestiones que nos darían un KPI_5 negativo.



ANALYTIC
HOUND