

Time 15m 1H 4H 1D 1W 1m ▾ 车 回

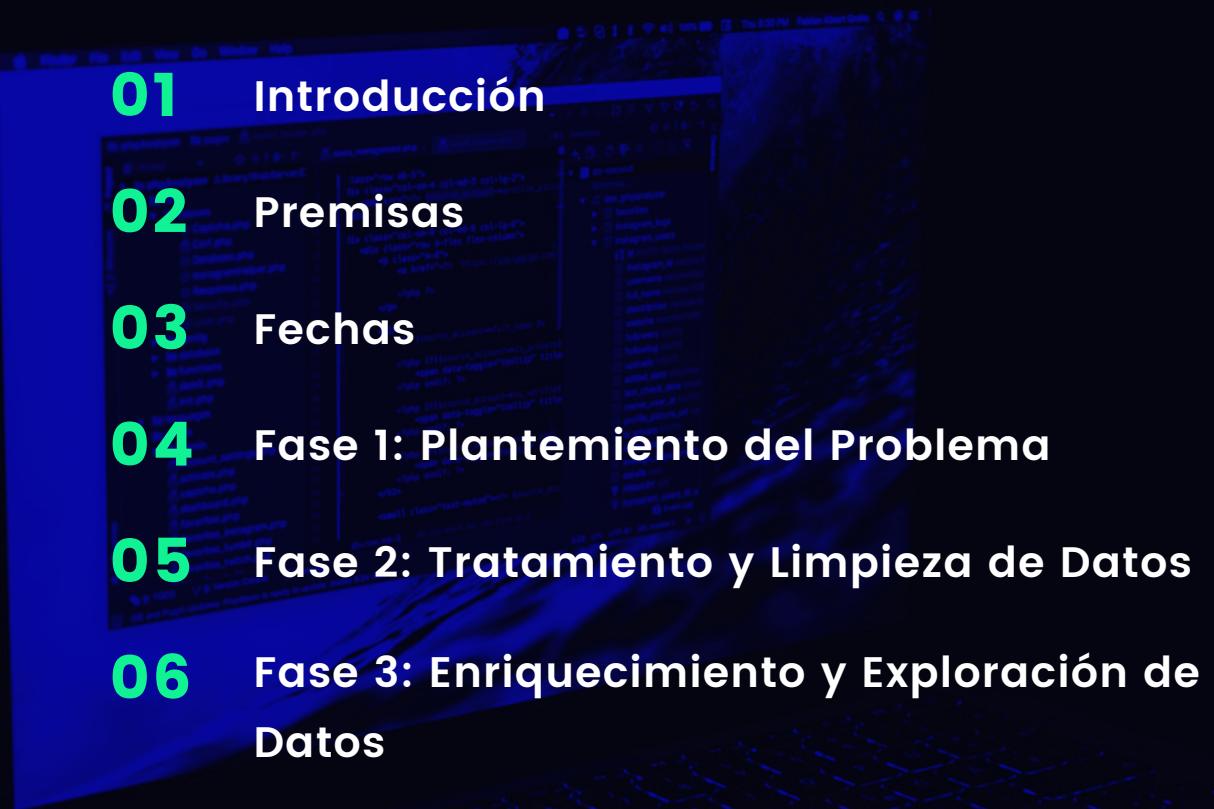
DT, 1 O 23.072 H 23.091 L 23.072 C 23.072

23.04126571
23.04616000
22.600

PROYECTO DATA ANALYTICS

MORE THAN LEARN, BECOME.

ÍNDICE

- 
- 01 Introducción**
 - 02 Premisas**
 - 03 Fechas**
 - 04 Fase 1: Plantamiento del Problema**
 - 05 Fase 2: Tratamiento y Limpieza de Datos**
 - 06 Fase 3: Enriquecimiento y Exploración de Datos**
 - 07 Fase 4: Machine Learning**
 - 08 Fase 5: Visualización**
 - 09 Fase 6: Puesta en Producción**
 - 10 Consejos**



PROYECTO FINAL

Introducción



Para evaluar los conocimientos adquiridos en las principales áreas vistas en el bootcamp de [Data Analytics](#) como lo son: Python, análisis y manipulación de datos, estadística y álgebra, POO, aprendizaje automático, dashboards/visualización, es necesario desarrollar un [proyecto final](#) por grupo.

El proyecto es de [carácter libre](#), buscando resolver un problema y esperando sea una idea que tenga sentido comercial. Para ello debéis elegir un tema, uno o varios conjuntos de datos, y realizar un [estudio detallado](#) sobre él.

Fases del Proyecto

Fase 1: Planteamiento del Problema

Fase 2: Tratamiento y Limpieza de Datos

Fase 3: Enriquecimiento y Exploración de Datos

Fase 4: Machine Learning

Fase 5: Visualización

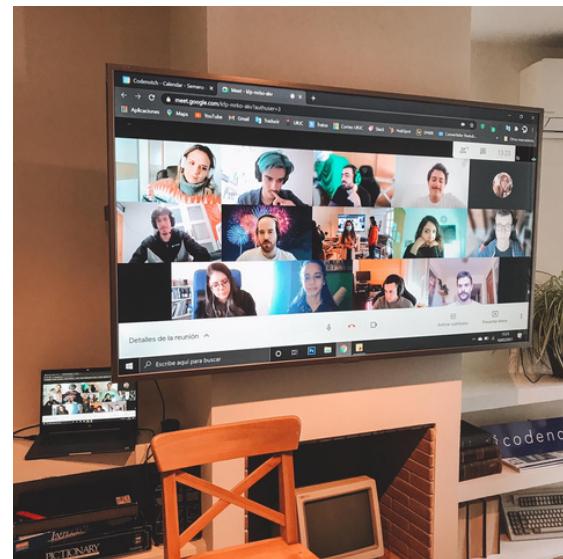
Fase 6: Puesta en Producción

Entrega del Proyecto



PROYECTO FINAL

Premisas



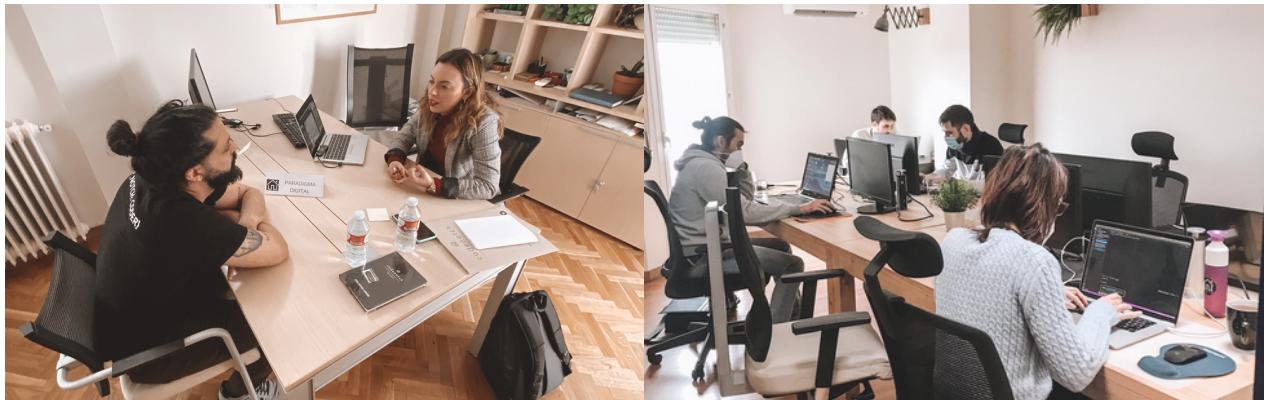
Para esta edición se tendrá un equipo con 5 integrantes.

- José Herrera será el tutor metodológico y de seguimiento de actividades, donde validará y asignará los objetivos/tareas diarias para la consecución del proyecto junto con la ayuda del resto del equipo docente.
- En caso de tener que tratar temas específicos, se recurrirá a Pablo Talavante, Félix Enriquez o Guillermo Yuste, en las fechas y horas especificadas en el [calendario escolar](#).
- Se realizará un acta de [compromiso](#) por cada grupo para dedicar un mínimo de [30 horas semanales](#) para desarrollar el proyecto (las correspondientes a las impartidas en clases).
- Se debe crear un repositorio en [GitHub](#) por equipo de trabajo de acuerdo con lo explicado en clases donde se maneje el código por este medio, con su respectivo README relacionado al proyecto.
- El presente documento recopila los requerimientos mínimos definidos por cada profesor y que deben ser contemplados en el desarrollo de todo el proyecto.
- La presentación final será en función a un [proyecto “idea de negocio”](#), y debe ser contado por el alumno a través del Data Storytelling.



PROYECTO FINAL

Fechas



- **Martes 28 de Noviembre:** Explicación de la metodología y procedimiento para el desarrollo del proyecto final, los **equipos de trabajo, metodología scrum**, premisas del proyecto por parte de **José Herrera**. A partir de esta fecha cada **grupo** debe ir trabajando en sus **ideas de proyectos**.
- **Martes 5 de Diciembre:** Presentación de **ideas** a los profesores del bootcamp. Se estudiará la **viabilidad** de los proyectos y entre todos se elegirá cual es el más adecuado para realizar.
- **Lunes 11 de Diciembre:** A partir de esta fecha los alumnos pueden comenzar a realizar el proyecto final.
- **Durante las 2 semanas posteriores**, se tendrán reuniones diarias con el tutor metodológico de máximo 15 min. Menos los lunes que se distribuirán las tareas semanales y por tanto la reunión será más larga.
- **Al final de cada fase** se deberá realizar un pequeño **entregable** para comprobar que dicha fase se está realizando de manera correcta.
- **A lo largo de estas dos semanas** se dispondrá de **6 horas al día** con la disponibilidad de los profesores para cualquier **duda o problema**. El profesor responsable de atender estas dudas está definido en el **calendario escolar**.
- **Viernes 22 de Diciembre:** Fecha presentación del proyecto.

FASE 1: DÍA 1

Planteamiento del Problema



En este apartado es esencial que el alumno intente acceder a los datos necesarios para el resto del proyecto. Por lo que es muy importante que obtenga la [información](#) necesaria de las [fuentes más apropiadas](#).

El método de obtención puede ser directamente con CSV, vía API REST o Web Scraping. Aunque todos estos métodos son válidos, si es posible, es preferible que sea vía [API REST](#).

Puntos Importantes

- ¿Cual es el problema?
- ¿Qué datos/APIs se van a usar?
- Como se planter el problema de Machine Learing
- Definir qué tipo de problema es el que se va a abordar (Supervisado/ No Supervisado, etc)

Entregables

- Documento con la descripción del problema.
- Diagrama de la arquitectura de datos.



FASE 2: DÍAS 2 & 3

Tratamiento & Limpieza de Datos

El conjunto de datos debe quedar adecuado para su uso. Esto quiere decir, que tenga nombres de columna representativos y usables (minúsculas, sin caracteres extraños) y que los [tipos de dato](#) sean los [adecuados](#) (fechas, números, . . . según aplique).

Según el objetivo perseguido por el proyecto, puede que sea interesante la eliminación de filas con muchos nulos o la imputación de estos.

Los datos están formados por varias tablas partidas por mes o año, pero con la misma estructura interna, la tabla final analizada deberá ser única, resultado de concatenar las anteriores.

Si los datos están en un formato difícil de manejar, como p.e., formato ancho, deberás convertirlo.

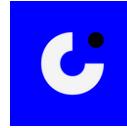
Puntos Importantes

- Establecer conexiones con las apis.
- Descargar los datos y organizarlos.
- Decidir qué BBDD utilizar (Mongo, MySQL, etc).
- Empezar a explorar estos datos.
- ¿Son datos en tiempo real? .
Obtener deltas.
- Dejar la BBDD o Dataset limpio.

Entregables

- Mostrar los datos limpios
(Dataset o BBDD)





FASE 3: DÍAS 4 & 5

Exploración y Enriquecimiento de Datos



Según cada caso, puede tener sentido incorporar información de [otras fuentes](#) que enriquezcan el conjunto de datos principal.

Esto, generalmente, implica [obtener](#), [tratar](#) y [limpiar](#) la [nueva fuente](#), y su cruce con los datos originales.

Se deberá plantear una serie de preguntas básicas sobre el dataset y combinar texto explicando los pasos con los gráficos obtenidos.

Algunas posibles preguntas son:

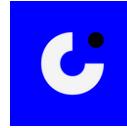
- ¿Cuál es la distribución de ciertas variables?
- ¿Cómo se relacionan estas variables entre sí?
- ¿Cómo evoluciona una variable a lo largo del tiempo?
- ¿Cómo se distribuyen geográficamente (en un mapa)?
- Puedes apoyar tus gráficas con tablas resumen extraídas de tus datos.

Puntos Importantes

- Descripción de los datos.
- Imputación de datos.
- Consideración de casos extremos (cuando faltan datos).
- Planteamiento del modelo de Machine Learning

Entregables

- Breve resumen de los pasos realizados para la exploración de datos.



FASE 4: DÍA 6

Machine Learning

Durante esta fase se realizará la preparación de datos para un [modelo de ML](#), comparando los diferentes [algoritmos](#) y decidiendo cual es el que se adapta mejor al objetivo que se desea conseguir.

También se realizará la [búsqueda](#) de [hiperparametros](#) y se preparará el modelo para realizar [inferencias](#).



Puntos Importantes

- Preparación de los datos para un modelo de ML.
- Feature engineering.
- Feature selection.
- Creación del modelo de ML.
- Comparativa de algoritmos.
- Búsqueda de hiper parámetros
- Ensamblado (opcional)
- Exportación del modelo.
- Preparar el modelo para hacer inferencias

Entregables

- Mostrar todo el proceso que ha realizado y el modelo en formato pickle.



FASE 5: DÍAS 7 & 8

Dashboards: Power BI & Streamlit

El análisis visual es el objetivo final en un proyecto de Data Analytics. Ayudándose en el [manejo de datos en distintos lenguajes](#), el proyecto deberá ser capaz de proporcionar información útil para la empresa, dando respuesta al problema planteado a través de [dashboards personalizados](#) y métricas clave.

Para esto, se profundizará el análisis de los datos mediante el reporte utilizando [Power BI](#), así como con la herramienta [Streamlit](#), que permite desarrollar aplicaciones web de visualización de datos.

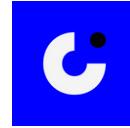
Puntos Importantes

- Bocetar los datos que se van a mostrar.
- Programar Power BI.
- Cargar los datos en Power BI service.
- Programar Streamlit.
- Mostrar inferencias de ML en Streamlit



Entregables

- Entrega de los cuadros de mandos tanto en PowerBI y Streamlit.



FASE 6: DÍA 9

Refinamiento del proyecto

En esta ultima fase se realizará un [repaso general](#) por parte de los miembros del [equipo](#) de desarrollo para comprobar que todos los [puntos solicitados](#) están realizados de [manera correcta](#).

Además en esta fase se debe crear una [presentación](#) (a través de un conjunto de [slides](#)) mostrando el problema planteado las soluciones propuestas y un resumen de los objetivos principales del proyecto. Esta presentación se deberá exponer el día de la [entrega del proyecto](#).

Finalmente se pretende que tanto el [dashboard en Power BI](#) como la [aplicación en Streamlite](#) sean pasadas a producción en esta última fase de desarrollo. Dichas aplicaciones se mostrarán el día de la [entrega del proyecto](#) justo [después](#) de la [presentación con slides](#).

Puntos Importantes

- Preparar presentación del proyecto.
- Poner el proyecto en producción.
- Cuadro de Mandos en PowerBI Service y StreamLit en producción.

Entregables

- Entrega de los cuadros de mandos tanto en PowerBI y Streamlit en producción.
- Creación de una presentación explicando el proyecto, sus objetivos y las técnicas utilizadas en su desarrollo.





CONSEJOS

Proyecto final del Bootcamp

Para comenzar el proyecto de [Data Analytics](#), se necesitará de una idea en la cual trabajar. Se recomienda hacer una [lluvia de ideas](#) sobre posibles áreas de interés. Durante este proceso, piensa en grande, ten una visión ambiciosa y no te limites, piensa lo más amplio y loco que puedas, no [censures tu creatividad](#). Una vez que tenga distintas propuestas, acordaremos de forma consensuada la idea más factible / interesante. Piensa en lo siguiente:

- ¿Qué tema te entusiasma lo suficiente como para disfrutar trabajando durante 2 a 3 semanas? ¿Es futbol, finanzas, psicología? ¿Qué tipo de datos existen en ese espacio? ¿Qué tipo de aplicaciones podrían crearse en torno a estos temas?
- ¿Tienes noción de algunos puntos débiles asociados con los productos y servicios que usas a diario y que puedan mejorarse?
- ¿Dónde quieras trabajar? ¿Qué tipo de aplicación mostraría tus habilidades a un posible empleador?
- ¿Hay productos que puedas diseñar para el bien social?
- ¿Qué tipo de herramientas te gustaría poner en práctica? ¿En qué tipo de proyectos podrías trabajar que te permitan utilizar estas herramientas? ¿Quiere mejorar tus conocimientos de SQL, AWS, Python, procesamiento de imágenes, Business Intelligence, etc?

Finalmente, se recomienda mantener una línea de tiempo con una sección de [Tareas pendientes](#), En progreso, Completado y Detenidas. Tener un [control diario](#) para ver lo que se logró el día anterior y establecer una meta para el nuevo día. Los objetivos y las métricas ayudarán a responsabilizarse y garantizar que realmente cumpla y [terminen el proyecto](#).



codenotch

MORE THAN
LEARN,
BECOME