

Proyecto Final

Sprint 1

2023



1 JUNIO

Analytix - Data consulting

Creado por:

EQUIPO SCRUM

- Matías Cejas
- David Soler
- Juan Arieta
- Ana María García
- Luis Bouchez

Contenido

Proyecto Yelps – Google Maps.....3

Entendimiento de la situación Actual3

Objetivo General.....3

Objetivos especificos3

Alcance del proyecto4

- Alcance Temporal:.....4
- Alcance del equipo:4
- Alcance Financiero:.....4
- Alcance Geográfico:.....5

Formulación de KPIs y Objetivos asociados.....5

KPI DESDE EL PUNTO DE VISTA DEL EMPRESARIO/NEGOCIO5

KPI DESDE EL PUNTO DE VISTA INVERSOR6

Repositorio en Github.....7

Solución Propuesta7

Stack Tecnológico7

Metodología de trabajo7

Entregables.....8

EQUIPO DE TRABAJO9

Cronograma del proyecto10

Proyecto Yelps – Google Maps

Entendimiento de la situación Actual

Como consultores de datos, nuestro enfoque se centra en analizar la información recopilada de plataformas de reseñas como Yelp y Google Maps. El objetivo principal de nuestro análisis es descubrir patrones y tendencias en las reseñas de los usuarios, con el fin de comprender su experiencia, necesidades y expectativas en relación a los diferentes servicios. Este análisis de datos proporcionará un análisis detallado a partir de los datos brindados para poder dar una asesoría y recomendaciones a posibles inversores que estén interesados en este proyecto, lo cual resulta sumamente valioso a la hora de tomar decisiones estratégicas y mejorar su imagen y reputación. Para lograrlo, aplicamos técnicas avanzadas de análisis de sentimientos y machine learning, con el propósito de brindar a nuestros clientes un modelo de recomendación que sea accesible y realista.

Objetivo General

Crear una plataforma integral que recopile, depure y disponibilice datos limpios y estructurados de reseñas de usuarios, realizando análisis exploratorio, construyendo un modelo de recomendación basado en análisis de sentimientos, un modelo predictivo de crecimiento de negocios y un dashboard interactivo para visualizar y explorar los resultados.

Objetivos específicos

1. **Garantizar la disponibilidad de datos limpios y estructurados:** Recopilando, depurando y disponibilizando la información relevante de plataformas de reseñas como Yelp y Google Maps. Esto asegurará la calidad y coherencia de los datos para su análisis y aplicación en el proyecto.
2. **Realizar un análisis exploratorio de las reseñas de usuarios:** Identificar patrones, tendencias y factores que influyen en la percepción de los usuarios sobre locales y servicios, utilizando las reseñas recopiladas de Yelp y Google Maps.
3. **Desarrollar un modelo de machine learning avanzado de recomendación:** Se construirá un modelo que utilice técnicas de análisis de sentimientos para clasificar

las reseñas y predecir las preferencias de los usuarios. Esto ofrecerá la posibilidad de optimizar la experiencia del cliente y desarrollar estrategias personalizadas

4. **Crear un modelo predictivo:** Se desarrollará como segundo modelo a presentar al empresario un modelo de aprendizaje automático (Machine Learning) que permita realizar predicciones sobre el crecimiento de los negocios, brindando así la capacidad de tomar decisiones acertadas en calidad de inversionista.
5. **Desarrollar un dashboard interactivo:** Crear una interfaz interactiva que visualice y explore los resultados del análisis de datos de las reseñas de usuarios. Proporcionar métricas, gráficos y estadísticas relevantes para facilitar la toma de decisiones estratégicas y la identificación de oportunidades de mejora en los servicios ofrecidos.

Alcance del proyecto

- **Alcance Temporal:**

Será de 4 semanas, con entregas semanales según SCRUM. Se utilizarán sprints de una semana para dividir las actividades. Las tres primeras semanas se completarán los hitos planificados, incluyendo recopilación de datos, análisis exploratorio y desarrollo de modelos de machine learning. La cuarta semana se dedicará a retoques y preparación de la presentación final.

- **Alcance del equipo:**

El equipo estará conformado por 2 ingenieros de datos, 2 especialistas en machine learning y 1 analista de datos. Los ingenieros se encargarán de la recopilación y depuración de datos, mientras que los especialistas desarrollarán modelos de machine learning. El analista realizará el análisis exploratorio de las reseñas. Cada miembro tendrá roles específicos para trabajar de manera colaborativa en todas las etapas. La comunicación y coordinación serán clave para integrar los componentes del proyecto.

- **Alcance Financiero:**

El alcance financiero del proyecto se enfoca en la utilización de herramientas gratuitas y recursos disponibles sin contar con un presupuesto asignado. El equipo, compuesto por 5 personas especializadas en ingeniería de datos, machine learning y data analytics, desarrollará el proyecto sin recibir remuneración, ya que forma parte del

proyecto sin fines de lucro. Se buscará aprovechar las opciones gratuitas para cumplir con los objetivos del proyecto de manera eficiente.

- **Alcance Geográfico:**

El proyecto se centrará en empresas del sector de la hotelería, restaurantes y otros negocios relacionados con el turismo y el ocio en los Estados Unidos.

Alcance de impacto:

El proyecto busca impactar tanto a empresarios como a clientes al proporcionar información valiosa sobre la percepción de la ronda de inversores y brindar recomendaciones personalizadas. Esto permitirá a los empresarios tomar decisiones estratégicas informadas y mejorar la experiencia del cliente. Con datos limpios, análisis exploratorio, modelos de machine learning y un dashboard interactivo, se optimizará la toma de decisiones, la experiencia del cliente y la relación entre ambas partes.

Formulación de KPIs y Objetivos asociados

KPI DESDE EL PUNTO DE VISTA DEL EMPRESARIO/NEGOCIO

Nombre_KPI	Descripción	Formula	Periodicidad	Objetivo
Tasa de satisfacción del cliente	Porcentaje de reseñas con una calificación alta (Entre 4 y 5)	$(\text{Cantidad de reseñas con calificación alta} / \text{Total de reseñas}) * 100$	Mensual	Garantizar una tasa de satisfacción al cliente del mínimo 75%
Tasa de respuesta del negocio	Porcentaje de reseñas a las que el negocio ha respondido en comparación con el total de reseñas recibidas.	$(\text{Reseñas respondidas} / \text{Total de reseñas}) * 100$	Mensual	Garantizar una tasa de respuesta al usuario de mínimo el 80%
Índice de interacción y nivel de engagement de las reseñas	Mide el nivel de interacción y participación de los usuarios con las reseñas, considerando los votos "useful", "funny" y "cool".	$(\text{Suma de votos "useful"} + \text{votos "funny"} + \text{votos "cool"}) / \text{Número total de reseñas}$	Trimestral	Aumentar el nivel de engagement del usuario en un 5%
Indicador de Crecimiento de Check-ins	Incremento en la cantidad de Check-ins en el periodo	$(\text{Cantidad de check-ins del periodo actual} - \text{Cantidad de check-ins del periodo anterior}) / \text{Cantidad de check-ins del periodo anterior} * 100$	Mensual	Aumentar el indicador de crecimiento de check-ins en un 5% mensual
Distribución de visitas por horario	Mide la proporción de visitas que se realizan en cada horario durante el día	Número de visitas que se realizan en cada horario durante el mes permitiendo identificar los momentos de mayor y menor afluencia de los clientes	Mensual	Lograr una distribución de visitas por horario que se acerque lo más posible a la meta establecida por el negocio

KPI DESDE EL PUNTO DE VISTA INVERSOR

Nombre_KPI	Descripción	Formula	Periodicidad	Objetivo
Valoración promedio de los negocios	Promedio de las calificaciones de los negocios basado en las reseñas recibidas.	Promedio de las calificaciones de las reseñas	Mensual	Encontrar negocios con valoraciones altas (4-5 estrellas) que indiquen una buena calidad y servicio.
Negocios con una tasa de respuesta positiva	Porcentaje de reseñas respondidas por los negocios	$(\text{Reseñas respondidas} / \text{Total de reseñas}) * 100$	Mensual	Identificar negocios que demuestren un compromiso mínimo del 80% al responder a las reseñas de sus clientes
Indicador de participación del cliente	Porcentaje de crecimiento de engagement de un periodo con respecto al anterior para los negocios	$((\text{Engagement actual} - \text{Engagement anterior}) / \text{Engagement anterior}) * 100$	Trimestral	Identificar establecimientos con alto engagement al interactuar y responder a las reseñas de clientes, mejorando la satisfacción y fortaleciendo la relación.
Indicador de Expansión de Clientela	Porcentaje de crecimiento de check-ins de un periodo con respecto a otro en los negocios	$(\text{Cantidad de check-ins del periodo actual} - \text{Cantidad de check-ins del periodo anterior}) / \text{Cantidad de check-ins del periodo anterior} * 100$	Mensual	Identificar establecimientos con un crecimiento positivo de clientes del X% durante un período determinado

Repositorio en Github

<https://github.com/Davoassassin27/PROYECTO-GRUPAL-GOOGLE-YELP>

Solución Propuesta

Stack Tecnológico

Tecnología	Descripción
Visual Studio Code	Entorno de desarrollo integrado para programación
Python/Pandas	Librería de Python para manipulación y análisis de datos
Python/Numpy	Librería de Python para cálculos numéricos
Python/Seaborn	Librería de visualización estadística en Python
Python/Matplotlib	Librería de Python para la generación de gráficos
Python/Scikit learn	Librería de Python para técnicas de machine learning
SQL	Lenguaje de consulta estructurado utilizado para manipular y consultar bases de datos
Tensor Flow	Framework de aprendizaje automático y deep learning
Google Cloud Plataform	Plataforma en la nube de Google para almacenamiento y procesamiento de datos
Power BI	Herramienta de visualización interactiva de datos
GitHub	Plataforma de control de versiones y colaboración en el desarrollo de software
Write	Herramienta de creación y edición de documentos
Canva	Herramienta de diseño gráfico en línea

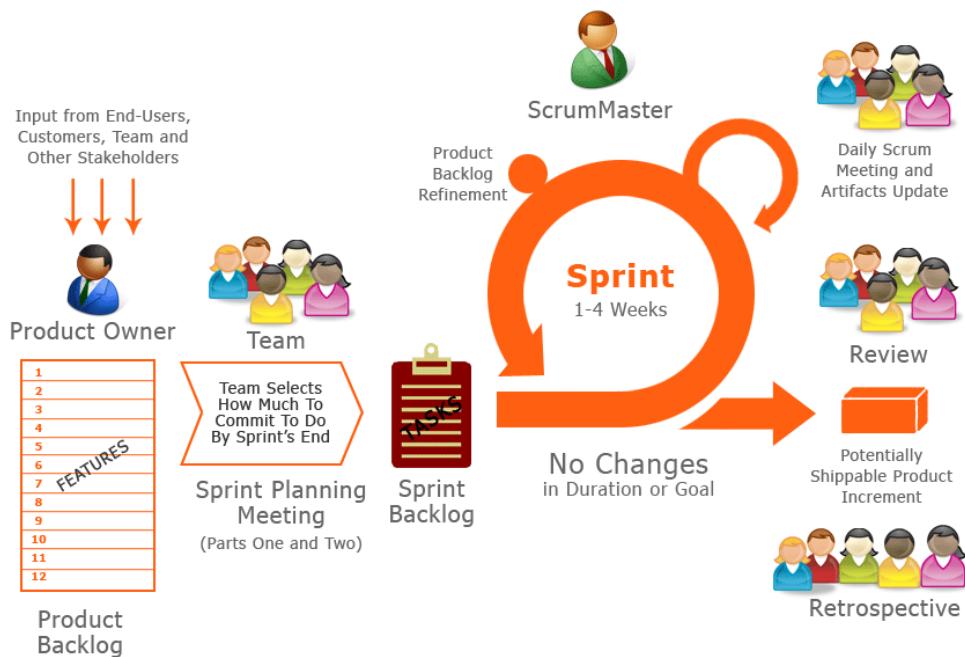
Metodología de trabajo

El ciclo inició con la reunión de planificación o Sprint Planning meeting donde se definieron los roles de los integrantes en el proyecto y se revisó el Backlog, sus sprints y entregables.

Cada sprint tiene una duración de 1 semana donde el equipo SCRUM trabajará en la creación de los entregables y cada día se realizarán reuniones diarias de seguimiento (Daily Standup) de 45 minutos con el Henry mentor, a fin de discutir el progreso diario y posibles inconvenientes.

Al finalizar cada sprint backlog se realizará una demo (sprint review meeting) en la que se hará una demostración de los entregables desarrollados, esperando una retroalimentación por parte del Henry Mentor.

El proyecto se realizará bajo el siguiente esquema:



Entregables

N° Sprint	Semana de ejecución	Entregable	Descripción
1	29-05-23 al 02-06-23	Propuesta del proyecto	<p>En esta semana se entrega la planificación del proyecto, el cual contiene:</p> <ul style="list-style-type: none"> Entendimiento de la situación actual. Objetivos del proyecto Alcance del proyecto Objetivos y KPIs asociados Creación de repositorio en Github Solución propuesta: Stack tecnológico escogido, Metodología de trabajo implementada, Diseño detallado de entregables, Roles y responsabilidades del equipo de trabajo, Cronograma del proyecto

2	05-06-23 al 09-06-23	Documentación del Proceso de Data Engineering	En esta semana se entrega la infraestructura del proyecto, el cual incluye: Diagrama ER detallado Diccionario de datos Workflow detallando tecnologías
3	12-06-23 al 16-06-23	Dashboard y Modelo de Machine learning	Esta semana se concretan los KPI y objetivos formulados y se crea un dashboard interactivo con el análisis de los datos que se han trabajado preparando un buen storytelling. También se crean los modelos de Machine Learning planificados desde el inicio.
4	19-06-23 al 23-06-23	Presentación de la demo	Se trabaja en las presentaciones, storytelling y se valida que todo funcione correctamente, completando la documentación.

EQUIPO DE TRABAJO

NOMBRE DEL INTEGRANTE	ROL EN EL PROYECTO	RESPONSABILIDADES
Matías Cejas (Líder del equipo Scrum)	Data Engineer	Realizar una selección cuidadosa del stack tecnológico, eligiendo las herramientas y tecnologías adecuadas que permitan cumplir con los requisitos y objetivos del proyecto, optimizando así la eficiencia y el rendimiento del sistema. Implementar medidas para garantizar la disponibilidad de datos limpios y estructurados, asegurando que la información utilizada esté completa, actualizada y libre de errores, lo que permitirá tomar decisiones informadas basadas en datos confiables y precisos.
Luis Ángel Bouchez Méndez (Integrante del equipo Scrum)		

David Andrés Soler (Integrante del equipo Scrum)	Data Scientist (Machine Learning)	Desarrollar un modelo de machine learning altamente preciso y eficiente que utilice técnicas de recomendación para predecir con exactitud las preferencias individuales de los usuarios, con el objetivo de ofrecer recomendaciones personalizadas y mejorar su experiencia.
Juan Arrieta (Integrante del equipo Scrum)		Crear un modelo predictivo avanzado utilizando técnicas de machine learning para realizar pronósticos precisos sobre el crecimiento futuro de los negocios, permitiendo tomar decisiones estratégicas basadas en datos y maximizar las oportunidades de crecimiento y rentabilidad.
Ana María García Arias (Integrante del equipo Scrum)	Data analyst	Desarrollar una interfaz interactiva que brinde la posibilidad de explorar de manera intuitiva los resultados del análisis de datos en las reseñas de los usuarios. La finalidad es facilitar la visualización y comprensión de los insights obtenidos, permitiendo una interacción dinámica con la información para una toma de decisiones más informada.

Cronograma del proyecto

Con el siguiente puede visualizar el estado del proyecto

<https://app-us2.wrike.com/workspace.htm?acc=5588043#folder/1129454736/tableview?filters=status%3Dactive&spaceId=917377012&viewId=163467489>

Nombre ↑	Estado	Fecha de inicio	Fecha de vencimie...
▼ PROYECTO FINAL	En curso	29/05/2023	23/06/2023
> Sprint 1 - Propuesta del proyecto	Completado	29/05/2023	02/06/2023
> Sprint 2 - Realización del proceso de Data Engineering	En curso	05/06/2023	09/06/2023
> Sprint 3 - Dashboard y Machine Learning	En espera	12/06/2023	16/06/2023
> Sprint 4 - Presentación del proyecto terminado	En espera	19/06/2023	23/06/2023
Elemento			