**Estructura de Presentación de Propuesta de Proyecto para Análisis de Datos**

**Grupo #8 Tester ACE** 

**Integrantes:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Identificación | Apellidos | Nombres | Correo | Teléfono |
|  | **Ruiz Acosta** | **Jorge Luis** | **Jorgeruizacosta117@gmail.com** |  |
|  | **De La Espriella Gándara** | **Eric David** | **ericddg12.23@gmail.com** |  |
|  | **Rojas Acevedo** | **Gissella** | **Rojasacevedo18@gmail.com** |  |
|  | **Martínez Contreras** | **Diana Karina** | **dianamcontreras22@gmail.com** |  |
|  |  |  |  |  |

**Título: Exploración de patrones colectivos en datos públicos de Strava mediante análisis estadístico y visualización interactiva**

**1. Introducción**

La propuesta busca aplicar técnicas de análisis de datos mediante la exploración y visualización de información proveniente de Strava, una plataforma digital que funciona como red social enfocada en el monitoreo del rendimiento deportivo. A diferencia de los dashboards nativos orientados a la experiencia individual, se plantea el análisis de un conjunto de datos públicos, anonimizados y estructurados, con el objetivo de identificar patrones y tendencias generales entre usuarios desde una perspectiva colectiva.

Para ello, se seguirá la metodología CRISP-DM, integrando análisis estadístico, visualización avanzada y narrativa basada en datos mediante herramientas como Python y Power BI. Esta aproximación permitirá derivar hallazgos interpretables y reproducibles que contribuyan a comprender el comportamiento deportivo desde una mirada analítica.

A futuro, este enfoque podrá escalarse mediante el desarrollo de paneles interactivos y modelos de predicción, fortaleciendo así su aplicabilidad como producto analítico en contextos reales.

**2. Definición del Problema**

Aunque Strava ofrece visualizaciones personalizadas del rendimiento deportivo para cada usuario, no proporciona herramientas para analizar de forma agregada el comportamiento colectivo de sus miembros. Esta limitación impide identificar patrones generales y tendencias comunes que podrían aportar valor a investigaciones deportivas, decisiones estratégicas o desarrollos futuros. Existe, por tanto, una oportunidad de aplicar técnicas de análisis de datos a registros públicos y estructurados de la plataforma para generar insights más amplios y comprensibles sobre la actividad física de la comunidad.

**3. Área del Proyecto**

Este proyecto se enmarca dentro del análisis de datos aplicado a entornos deportivos, con énfasis en la exploración estadística, visualización de información y construcción de narrativas basadas en datos. Utiliza herramientas como Python y Power BI, y se apoya en la metodología CRISP-DM para estructurar cada fase del análisis de manera técnica y replicable.

**Justificación**

En un entorno donde las plataformas digitales generan grandes volúmenes de información, como es el caso de Strava, surge la oportunidad de aplicar técnicas de análisis de datos para obtener hallazgos significativos a partir de datos públicos y anonimizados. Aunque Strava ofrece paneles individuales para sus usuarios, no proporciona una visión agregada del comportamiento colectivo. Este proyecto busca llenar ese vacío mediante un enfoque analítico que permita identificar tendencias generales, patrones de rendimiento y posibles relaciones entre variables deportivas, aportando valor desde una perspectiva técnica, exploratoria y reproducible. Además, sienta las bases para futuros desarrollos como modelos predictivos o dashboards interactivos más avanzados.

**4. Objetivo General**

Aplicar análisis estadístico y visual sobre datos públicos de Strava para descubrir patrones colectivos y facilitar su visualización mediante dashboards interactivos.

**5. Objetivos Específicos**

1. Examinar la estructura, calidad y variables disponibles en el dataset de Strava.
2. Aplicar técnicas estadísticas para identificar patrones y tendencias en la actividad física registrada.
3. Desarrollar visualizaciones claras y narrativas de datos que destaquen los hallazgos clave.
4. Interpretar los resultados obtenidos para comprender el comportamiento colectivo de los usuarios.
5. Proponer desarrollos futuros, incluyendo modelos predictivos y dashboards interactivos para el análisis comparativo de datos.

**6. Alcance del Proyecto**

Este proyecto se enfoca exclusivamente en el análisis exploratorio y estadístico de datos deportivos públicos provenientes de Strava. A través de la metodología CRISP-DM, se realizará una comprensión del contexto, limpieza de datos, análisis descriptivo, visualización de patrones y generación de hallazgos interpretables.

El desarrollo se realizará con herramientas como Python y Power BI, limitándose a un entorno académico y de aprendizaje.

No se contempla la creación de una aplicación funcional o interfaz de usuario interactiva, aunque se dejarán recomendaciones para posibles desarrollos futuros como paneles interactivos avanzados o modelos predictivos.

El análisis estará centrado en variables relacionadas con el rendimiento físico (como duración, velocidad, distancia y frecuencia), siempre a partir de datos públicos y anonimizados, respetando la privacidad de los usuarios.

**7. Sugerencias para desarrollo futuro**

Aunque el presente proyecto se limita al análisis exploratorio y visualización de datos públicos de Strava con fines académicos, se identifica un alto potencial de evolución hacia una solución analítica más robusta y aplicable en escenarios reales. A futuro, se sugiere:

* Desarrollar modelos predictivos que permitan estimar métricas de rendimiento deportivo a partir de variables como distancia, frecuencia y duración, utilizando técnicas de machine learning.
* Implementar dashboards interactivos que faciliten la exploración dinámica de tendencias y comparaciones entre usuarios, con filtros personalizables y visualizaciones optimizadas.
* Construir una aplicación web o móvil que permita integrar estos paneles, modelos y análisis para el uso por parte de comunidades deportivas, clubes o entrenadores, con énfasis en la visualización colectiva y no individual.
* Incorporar análisis longitudinales para estudiar la evolución del rendimiento deportivo a lo largo del tiempo.
* Ampliar el alcance del dataset con fuentes complementarias, manteniendo siempre la anonimización y la protección de datos personales.

Estas sugerencias permitirían transformar el proyecto en una herramienta de análisis deportivo avanzada, fomentando la toma de decisiones basadas en evidencia y la generación de valor para distintas audiencias en el ecosistema fitness y deportivo.

**8. Metodología (Basada en CRISP-DM)**

* ***Enfoque General:*** Explicar que se usará CRISP-DM, un estándar para análisis de datos, adaptado a un nivel básico para construir el dashboard y aplicar storytelling. Las fases son:
  + ***1. Comprensión del Negocio:*** Definir el objetivo (ejemplo: entender ventas) y cómo el dashboard lo resolverá. Establecer qué historia se quiere contar.
  + ***2. Comprensión de los Datos:*** Explorar los datos del área (ejemplo: revisar columnas como fechas y ventas) para conocer su estructura y calidad.
  + ***3. Preparación de los Datos:*** Limpiar y organizar los datos (eliminar errores, calcular totales) para que sean aptos para el dashboard.
  + ***4. Modelado:*** Aplicar análisis simple (como promedios o tendencias) para encontrar insights que se visualizarán y narrarán.
  + ***5. Evaluación:*** Revisar si el dashboard refleja los objetivos y si el storytelling es claro y útil.
  + ***6. Despliegue:*** Finalizar el dashboard y presentarlo, explicando su uso práctico (ejemplo: para planificar inventarios).
* ***Herramientas:*** Usar software como Excel, Google Sheets, LookerStudio, Tableau o Power BI para análisis y visualización.
* ***Storytelling:*** Diseñar el dashboard para guiar al usuario a través de una narrativa (ejemplo: mostrar primero el problema, luego el patrón y finalmente la solución).

**9. Plan de Trabajo**

* ***Cronograma:*** Alinear las fases de CRISP-DM con un calendario:
  + ***Semana 1:*** Comprensión del negocio y datos (fases 1 y 2).
  + ***Semana 2:*** Preparación de datos y análisis inicial (fases 3 y 4).
  + ***Semana 3***: Diseño del dashboard y evaluación (fases 4 y 5).
  + ***Semana 4:*** Despliegue y presentación final (fase 6).
* ***Entregables:*** Datos limpios, análisis preliminar, dashboard terminado.

**10. Recursos Necesarios**

* ***Recursos Técnicos:*** Software (Power BI, Excel), computador.
* ***Datos:*** Conjunto de datos del área elegido, accesible y adecuado.
* ***Apoyo:*** Instructor o compañero.

**11. Evaluación y Métricas de Éxito**

* ***Criterios de Éxito:*** Basados en la fase 5 de CRISP-DM:
  + El dashboard muestra datos claros y responde al problema.
  + El storytelling es comprensible y atractivo.
* ***Métricas:*** Precisión de los datos, facilidad de uso del dashboard, feedback del curso.

**12. Conclusión**

* ***Resumen:*** Recapitular el uso de CRISP-DM para crear un dashboard útil en el área elegida, con storytelling efectivo.