

Momento de Retroalimentación:
Análisis del contexto y la normatividad.
(Portafolio Análisis - Individual)

Instituto Tecnológico y de Estudios Superiores de Monterrey
Escuela de Ingeniería y Ciencias
Ingeniería en Ciencia de Datos y Matemáticas

Francisco Castorena Salazar A00827756

Inteligencia artificial avanzada para la ciencia de datos I (Gpo 101)

Monterrey, Nuevo León, 29 de Agosto del 2023

1. Antecedentes sobre el proyecto de análisis de datos tratado en este documento

El proyecto de análisis de datos que se abordará en este documento fue realizado como parte del curso de Google Data Analytics Certificate, usando el dataset de Cyclistic Bike Share, los datos se pueden encontrar en el siguiente link: <https://www.kaggle.com/datasets/evangower/cyclistic-bike-share> el cual se basa en información ficticia derivada de datos reales de la compañía de bicicletas compartidas Divvy Bikes, es un ejercicio que requiere el cumplimiento de normativas y regulaciones relacionadas con la privacidad, la seguridad de datos y el uso ético de la información. En este documento, se justifica el uso de estos datos teniendo en cuenta la normativa asociada.

2. Normativa Asociada a los Datos

El dataset utilizado en el proyecto Google Data Analytics Capstone Project proviene de Cyclistic, una empresa ficticia basada en datos reales de la compañía Divvy Bikes. Para asegurar el cumplimiento de la normativa, se consideraron las políticas de privacidad y las leyes de protección de datos de Estados Unidos y México. La normativa aplicable incluye:

Regulaciones de Privacidad de los Estados Unidos: Se tomaron en cuenta las regulaciones federales, como la Ley de Protección de la Privacidad en Línea de los Niños (COPPA) y la Ley de Protección de la Privacidad del Consumidor de California (CCPA), para asegurarse de que la recopilación y el uso de datos se alinearan con estas leyes.

Leyes de Protección de Datos de México: Para garantizar el cumplimiento con la normativa mexicana, se siguieron las disposiciones de la Ley Federal de Protección de Datos Personales en Posesión de los Particulares (LFPDPPP).

El dataset de Cyclistic Divvy Data se utilizó para analizar patrones de uso de bicicletas compartidas en México. Para evitar violar la normativa, se aplicaron técnicas de anonimización en los datos, eliminando información personal identificable. Se mantuvo el enfoque en patrones generales y tendencias en lugar de detalles individuales.

La utilización responsable de los datos se basa en documentos oficiales como las políticas de privacidad de Divvy Bikes y las regulaciones legales mencionadas anteriormente.

3. Medio utilizado para el análisis de datos acorde a la norma establecida

Para el análisis del proyecto Google Data Analytics Capstone Project con el dataset de Cylcistic Divvy Data, se optó por utilizar el lenguaje de programación R debido a su robusta capacidad para análisis de datos y visualización. El uso de estos datos se llevó a cabo con especial atención a no violar la normativa establecida en términos de privacidad y protección de datos.

El cumplimiento con la normativa se respaldó mediante la revisión detallada de documentos oficiales relacionados con las políticas de privacidad y protección de datos de Divvy Bikes, así como con las leyes y regulaciones pertinentes en los Estados Unidos y México. Estos documentos fueron consultados y considerados como guías para el tratamiento ético de los datos:

1. Política de Privacidad de Divvy Bikes: Este documento proporcionó información sobre cómo se recopilan, almacenan y utilizan los datos de los usuarios. Se aseguró que cualquier análisis realizado estuviera dentro de los límites establecidos por estas políticas. (*Fuente: Divvy Bikes, Política de Privacidad*)

2. Ley Federal de Protección de Datos Personales en Posesión de los Particulares (LFPDPPP): Esta ley mexicana regula la recopilación y el tratamiento de datos personales. Se aseguró que el análisis no involucrara el uso indebido de datos personales y que se respetaran los principios de finalidad y consentimiento. (*Fuente: Diario Oficial de la Federación, LFPDPPP*)

3. Leyes de Privacidad Estadounidenses (COPPA y CCPA): Se consideraron las leyes de privacidad de los Estados Unidos para garantizar el tratamiento adecuado de datos, especialmente si involucraban a usuarios menores de edad o si los datos estaban siendo utilizados en California. (*Fuentes: Comisión Federal de Comercio, Ley de Protección de la Privacidad en Línea de los Niños (COPPA); Gobierno de California, Ley de Protección de la Privacidad del Consumidor de California (CCPA)*)

En el análisis con R, se adoptaron prácticas para proteger la privacidad y evitar la identificación de individuos. Se utilizaron técnicas de agregación y anonimización en los datos antes de cualquier análisis específico. Además, se generaron visualizaciones y resultados agregados que no permitieran la identificación de usuarios individuales.

4. Verificando que mi solución no incurra en ningún sesgo ético

La solución de análisis desarrollada utilizando el lenguaje de programación R se diseñó y ejecutó con el propósito de cumplir con la normativa de la industria y evitar cualquier forma de sesgo ético. Para lograrlo, se implementaron medidas concretas:

4.1. Cumplimiento Normativo

- Privacidad de los Usuarios: La solución garantiza el anonimato de los usuarios al utilizar técnicas de agregación y eliminación de información personal identificable. Esto está en línea con las regulaciones de privacidad tanto en los Estados Unidos como en México, como la Ley Federal de Protección de Datos Personales en Posesión de los Particulares (LFPDPPP) y las leyes de privacidad de los Estados Unidos.

- Consentimiento Informado: La herramienta de análisis no utiliza información que no haya sido previamente autorizada para su uso. Los datos se manipulan con el consentimiento implícito de acuerdo con las políticas de privacidad establecidas en el dataset original.

4.2. Prevención de Sesgos Éticos

- Diseño de Análisis Neutral: La solución se diseñó para enfocarse en patrones generales y tendencias en lugar de detalles individuales. Esto reduce la posibilidad de sesgos basados en características personales.

- Validación de Resultados: Se llevaron a cabo pruebas de robustez y sensibilidad para identificar posibles sesgos en los resultados obtenidos. Si se hubiera detectado algún sesgo, se habrían tomado medidas correctivas.

En resumen, la solución de análisis desarrollada cumple con la normativa de la industria mediante el uso responsable de los datos y la aplicación de técnicas de privacidad. Asimismo, se aborda la ética al evitar sesgos a través del diseño de análisis neutral y la validación de resultados. De esta manera, la herramienta asegura que los resultados sean confiables, éticos y cumplan con las regulaciones pertinentes.

5. Escenarios bajo los cuales el análisis/herramienta podría incurrir en una falta ética

El mal uso o la negligencia en el manejo de la herramienta de análisis pueden dar lugar a varias faltas éticas que podrían comprometer la privacidad de los usuarios y distorsionar los resultados. Algunos escenarios potenciales incluyen:

1. **Desanonimización Intencionada:** Si un usuario con malicia intenta desanonimizar los datos agregados para identificar a individuos específicos, se estaría violando la privacidad de los usuarios y poniendo en riesgo su información personal.
2. **Discriminación Injusta:** Si los resultados del análisis se utilizan para discriminar a ciertos grupos de usuarios en función de características como género, edad o ubicación, se estaría perpetuando sesgos y causando daño injusto a ciertos segmentos de la población.
3. **Discriminación Injusta:** Si los resultados del análisis se utilizan para discriminar a ciertos grupos de usuarios en función de características como género, edad o ubicación, se estaría perpetuando sesgos y causando daño injusto a ciertos segmentos de la población.
4. **Uso para Persecución:** Si la herramienta se utiliza para rastrear o acosar a usuarios individuales, ya sea por razones personales o comerciales, se estaría violando su privacidad y generando un ambiente perjudicial.
5. **Falta de Transparencia:** Si los resultados del análisis se presentan de manera engañosa o se interpretan de forma selectiva para respaldar agendas particulares sin proporcionar contexto completo, se estaría manipulando la información y perjudicando la confianza en los resultados.
6. **Incumplimiento de Consentimiento:** Si la herramienta se utiliza para analizar datos que no cuentan con el consentimiento explícito o implícito de los usuarios, se estaría infringiendo la normativa de privacidad y comprometiendo la ética del análisis.
7. **Falta de Responsabilidad:** Si no se toman medidas adecuadas para proteger la seguridad de los datos utilizados en la herramienta, como no implementar medidas de seguridad adecuadas, podría haber filtraciones que afecten negativamente a los usuarios y violen su confianza.

6. Link al análisis de datos referido en este documento

El análisis de datos en formato HTML puede ser encontrado en el siguiente link https://drive.google.com/file/d/11ni3N_R08Z1vRUgHabRnb28xCwAB3bEt/view?usp=drive_link