**Proyecto de Análisis de Datos en Profundidad de BiciMAD en Madrid**

**Resumen:**

Este proyecto tiene como objetivo realizar un análisis en profundidad del sistema de bicicletas públicas compartidas en Madrid, conocido como BiciMAD. Gestionado por el Ayuntamiento de Madrid, BiciMAD promueve la movilidad sostenible y el transporte público en la ciudad. Utilizaremos técnicas de machine learning para comprender y prever patrones de uso, así como para caracterizar demográficamente a los usuarios.

Para un gestor de BiciMad, serían interesantes predecir los siguientes datos de negocio:

- \*\*Demanda de bicicletas\*\*: Estimar la cantidad de usuarios que querrán alquilar bicicletas en diferentes momentos del día y semana. Esto ayudaría a planificar adecuadamente el inventario y evitar tener bicicletas sin uso o insuficientes para satisfacer la demanda.

- \*\*Ubicación de las estaciones de bicicletas\*\*: Identificar las áreas con mayor demanda y donde se necesiten más estaciones. Esto podría implicar analizar datos geográficos y tráfico de usuarios.

- \*\*Tiempo de espera promedio\*\*: Predecir el tiempo que los usuarios están esperando para obtener una bicicleta cuando llegan a una estación. Esto permitiría optimizar la gestión de la flota y mejorar la experiencia del usuario.

- \*\*Costos operativos\*\*: Estimar los costos asociados con la operación diaria, incluyendo mantenimiento de bicicletas, costos de personal, alquiler de estaciones, entre otros. Una predicción precisa de estos costos es crucial para mantener la rentabilidad del negocio.

- \*\*Factores climáticos\*\*: Predictar el impacto del clima en el uso de las bicicletas, como la probabilidad de lluvia o temperaturas extremas, que pueden afectar la disponibilidad y la demanda de servicios de alquiler.

- \*\*Eventos y festividades\*\*: Analizar la influencia de eventos locales o nacionales en la demanda de alquiler de bicicletas, ya que estas ocasiones suelen aumentar el uso de este medio de transporte.

- \*\*Competencia y cambios en el mercado\*\*: Monitorear la competencia y adaptarse a cambios en el mercado, como nuevas empresas que ofrecen servicios similares o cambios en las políticas gubernamentales relacionadas con la movilidad sostenible.

**1. Definición del Problema:**

* **Objetivos del Proyecto:**
  + Identificar y comprender patrones de uso de bicicletas en el sistema BiciMAD.
  + Prever la demanda de bicicletas para el año 2019 mediante un modelo predictivo.
  + Caracterizar demográficamente a los usuarios para comprender comportamientos específicos.
* **Contexto del Problema:**
  + El Ayuntamiento de Madrid gestiona BiciMAD para fomentar la movilidad sostenible.
  + Limitaciones incluyen disponibilidad de bicicletas, estaciones y condiciones climáticas.

**2. Entendimiento de los Datos:**

* **Recolección de Datos:**
  + Obtener datos históricos de uso de BiciMAD.
  + Incluir información sobre estaciones, usuarios y condiciones climáticas.
* **Análisis Exploratorio de Datos (EDA):**
  + Explorar distribuciones, correlaciones y outliers.
  + Identificar patrones diarios, estacionales y anuales.
* **Limpieza de Datos:**
  + Manejar valores nulos y atípicos.
  + Asegurar consistencia y calidad en los datos.

**3. Definición de Objetivos de Machine Learning:**

* **Selección de Tipo de Problema:**
  + Problema de regresión para prever la demanda de bicicletas.
  + Problema de clasificación para la caracterización demográfica de usuarios.
* **Selección de Métricas de Evaluación:**
  + MSE (Mean Squared Error) para modelos de regresión.
  + Accuracy y F1-score para modelos de clasificación.

**4. Preparación de Datos:**

* **División de Datos:**
  + Separar datos en conjuntos de entrenamiento y prueba.
* **Normalización/Estandarización:**
  + Ajustar escalas y distribuciones de variables numéricas.
* **Codificación de Características:**
  + Convertir variables categóricas en formato adecuado para modelos.

**5. Desarrollo de Modelos:**

* **Selección de Modelos:**
  + Iniciar con modelos simples como regresión lineal y logistic regression.
  + Explorar modelos más complejos como Random Forest y redes neuronales.
* **Divulgación de Hiperparámetros:**
  + Ajustar parámetros para optimizar rendimiento de los modelos.
* **Entrenamiento del Modelo:**
  + Utilizar datos de entrenamiento para enseñar al modelo.

**6. Evaluación del Modelo:**

* **Pruebas con Datos de Prueba:**
  + Evaluar rendimiento del modelo en datos no vistos.
  + Utilizar validación cruzada para asegurar robustez.
* **Ajuste de Modelo:**
  + Realizar ajustes según el rendimiento y análisis de errores.

**7. Interpretación de Resultados:**

* **Análisis de Errores:**
  + Comprender y corregir los errores del modelo.
* **Interpretación de Características:**
  + Analizar la importancia de variables en la toma de decisiones del modelo.

**8. Despliegue del Modelo:**

* **Implementación en Producción:**
  + Integrar el modelo en entornos de producción.
* **Monitoreo Continuo:**
  + Supervisar rendimiento del modelo y realizar actualizaciones según sea necesario.

**9. Documentación:**

* **Documentar Código:**
  + Asegurar documentación clara y detallada del código.
* **Documentar Resultados:**
  + Detallar hallazgos, decisiones y limitaciones del proyecto.

**10. Iteración:**

* **Mejora Continua:**
  + Basándose en la retroalimentación, mejorar el modelo y el proceso.

**11. Comunicación de Resultados:**

* **Informe Final:**
  + Presentar resultados, conclusiones y recomendaciones.
  + Incluir visualizaciones y explicaciones claras para audiencia no técnica.

**12. Gestión de Proyectos:**

* **Planificación y Seguimiento:**
  + Mantener un enfoque organizado y gestionar el tiempo eficientemente.
  + Utilizar metodologías ágiles para adaptarse a cambios y asegurar entregas puntuales.

**Objetivos Específicos del Proyecto:**

**1. Análisis Exploratorio de Datos (EDA):**

* Comprender la evolución del uso de bicicletas a lo largo de los años.
* Identificar patrones estacionales y tendencias en el uso del servicio de BiciMAD.

**2. Construcción de un Modelo Predictivo:**

* Desarrollar un modelo predictivo que estime la demanda de bicicletas para el año 2019.
* Utilizar características como el tiempo, patrones históricos y datos demográficos para mejorar la precisión del modelo.

**3. Caracterización Demográfica de Usuarios:**

* Categorizar a los usuarios por generación para comprender las preferencias y comportamientos de diferentes grupos demográficos.

**4. Integración con Git y Creación de Versiones:**

* Implementar la gestión de código colaborativo utilizando Git (GitHub).
* Asegurar que el código del proyecto esté versionado para facilitar la colaboración y el seguimiento de cambios.

**5. Creación de Aplicación Interactiva:**

* Convertir el cuaderno interactivo en una aplicación interactiva utilizando las capacidades de Jupyter.
* Proporcionar una interfaz que permita a los usuarios explorar y entender de manera intuitiva los resultados del análisis.

**Hitos del Proyecto (Orientativo):**

**1. Semana 1-2: Configuración y Exploración de Datos Iniciales:**

* Configurar el entorno de trabajo, incluyendo Visual Studio Code, Jupyter Notebook, Git y MySQL
* Importar y explorar los datos de BiciMAD para comprender su estructura y calidad.

**2. Semana 3-4: Análisis Exploratorio de Datos:**

* Realizar el análisis exploratorio de datos para identificar patrones de uso, fluctuaciones estacionales y posibles correlaciones.

**3. Semana 5-6: Construcción del Modelo Predictivo:**

* Seleccionar y preparar las características relevantes para la construcción del modelo.
* Implementar modelos de regresión lineal y polinómica y evaluar su rendimiento.

**4. Semana 7-8: Análisis Demográfico y Visualizaciones:**

* Categorizar a los usuarios por generación y realizar análisis demográficos.
* Crear visualizaciones interactivas con Plotly y Noteable para representar los resultados clave.

**5. Semana 9-10: Integración con Git y Creación de Versiones:**

* Establecer un repositorio en GitHub para gestionar el código de manera colaborativa.
* Garantizar la creación de versiones adecuadas para un seguimiento efectivo.

**6. Semana 11-12: Creación de Aplicación Interactiva:**

* Utilizar las capacidades de Streamlit/Render para convertir el cuaderno interactivo en una aplicación.
* Refinar la interfaz de usuario y asegurar que la aplicación sea fácil de usar.

**7. Semana 13: Documentación y Presentación:**

* Documentar el proceso, decisiones clave y resultados obtenidos.
* Preparar una presentación que destaque los hallazgos y la utilidad de la aplicación interactiva.Principio del formulario

