**Lineamientos para escritura de Ensayo:**

Ejemplo: Propuesta de Proyecto Final de Ciencia de Datos: Análisis de Problemas Socioeconómicos en Venezuela

**Objetivo General:**

El objetivo principal de este proyecto es que los estudiantes apliquen sus conocimientos y habilidades en el lenguaje de programación R para analizar un problema socioeconómico relevante para Venezuela. Los estudiantes deberán seguir las fases estándar del ciclo de vida de proyectos de ciencia de datos: importar, ordenar, transformar-limpiar, visualizar, modelar y comunicar.

**Fase 1: Selección de la Problemátic**a

**1. Identificación de Temas Relevantes:**

- Los estudiantes deben identificar un tema socioeconómico relevante para Venezuela que les interese personalmente.

- Algunos ejemplos pueden incluir:

- Análisis de la inflación y su impacto en el consumo

- Evaluación del desempleo y las tendencias laborales

- Estudio de la distribución de ingresos y pobreza

- Análisis de la calidad de vida en diferentes regiones del país

- Investigación sobre el acceso a servicios básicos (agua, electricidad, salud)

- Los estudiantes deben justificar por qué este tema les interesa y cómo creen que puede contribuir al campo de la ciencia de datos.

2. **Definición de la Pregunta de Investigación:**

- Basándose en el tema seleccionado, los estudiantes deben formular una pregunta de investigación clara y específica.

- La pregunta debe ser susceptible de análisis mediante técnicas estadísticas y visualización de datos.

**3. Evaluar posibles Fuentes de Datos:**

- Junto con el profesor se evaluará la viabilidad de obtener y recolectar periódicamente datos que sirvan para el proyecto. Se puede evaluar que es factible o no realizar la investigación.

**En este momento debería empezar el AnteProyecto juntando todas las ideas de cada participante de su correspondiente equipo**

**4. Selección de Fuentes de Datos:**

- Los estudiantes deben identificar y seleccionar fuentes de datos relevantes para su proyecto.

- Las fuentes pueden incluir:

- Estadísticas oficiales del INE (Instituto Nacional de Estadística)

- Informes económicos del Banco Mundial

- Datos de encuestas sociales

- Series temporales económicas

- Datos abiertos disponibles en plataformas como Kaggle o Data.gov

- Los estudiantes deben presentar una propuesta detallada de las fuentes de datos que planean utilizar, incluyendo un análisis de la calidad y relevancia de los mismos.

- La selección final de las fuentes de datos será revisada por el profesor para asegurar su adecuación al proyecto.

**Lo que sigue es lo que sí trabajaremos en el Proyecto**

**Fase 2: Importar y Ordenar Datos**

1**. Importar Datos:**

- Los estudiantes deben aprender a importar datos desde diferentes formatos (CSV, Excel, JSON, bases de datos SQL) utilizando R.

**2. Ordenamiento y Exploración Inicial:**

- Los estudiantes deben explorar los datos importados para entender su estructura y contenido.

- Deben identificar variables relevantes, tipos de datos y posibles problemas de calidad (valores faltantes, outliers, etc.).

**Fase 3: Transformar y Limpiar Datos**

**1. Transformación de Datos:**

- Los estudiantes deben aprender a transformar los datos para prepararlos para el análisis.

- Esto puede incluir la creación de nuevas variables, la recodificación de variables existentes o la aplicación de funciones matemáticas.

**2. Limpieza de Datos:**

- Los estudiantes deben identificar y manejar valores faltantes, outliers y otros problemas de calidad en los datos.

- Se les recomienda utilizar paquetes como `dplyr` para realizar operaciones de transformación y limpieza eficientes.

**Fase 4: Visualización de Datos**

**1. Creación de Gráficos y Tablas:**

- Los estudiantes deben aprender a crear gráficos y tablas que ayuden a visualizar los datos y responder la pregunta de investigación.

- Se les recomienda utilizar paquetes como `ggplot2`, plotly y otros paquetes que sean vistos en clase, para crear gráficos estéticamente atractivos y informativos.

**2. Interpretación de Visualizaciones:**

- Los estudiantes deben interpretar las visualizaciones creadas y relacionarlas con la pregunta de investigación.

- Deben ser capaces de comunicar claramente los hallazgos visuales en un contexto narrativo.

**Fase 5: Modelado de Datos: (no es obligatorio)**

1. **Selección de Modelos Estadísticos** - Los estudiantes deben seleccionar y aplicar modelos estadísticos apropiados para responder la pregunta de investigación.

- Esto puede incluir regresiones lineales, análisis de variancia (ANOVA), series temporales o otros métodos según sea necesario.

**2. Interpretación de Modelos:**

- Los estudiantes deben interpretar los resultados obtenidos de los modelos y relacionarlos con la pregunta de investigación.

- Deben ser capaces de comunicar claramente las implicaciones de los hallazgos estadísticos en un contexto narrativo.

**Fase 6: Comunicación de Resultados** **una de las fases más importantes en lo que estamos haciendo**

**1. Creación de Informes, Dashboard o Aplicación Web:**

- Los estudiantes deben aprender a crear informes que presenten sus análisis y resultados.

- Se les recomienda utilizar herramientas como qmd para combinar código, texto y visualizaciones en un solo documento. De esto hablaremos con detalle a lo largo del curso.

**2. Presentación Oral:**

- Los estudiantes deben preparar una presentación oral breve (10-15 minutos) que resuma su proyecto, incluyendo la pregunta de investigación, los métodos utilizados, los resultados obtenidos y las conclusiones.

- La presentación debe ser clara, concisa y visualmente atractiva.

**3. Presentación en un Afiche:**

- Luego conversaremos sobre este punto