
Preferencias de Espacios de Estudio en la Universidad: Un Análisis Basado en Datos

Presentado por:
Erick José Valenzuela López

Jhon Fred Martínez Utria

1. Introducción

El siguiente informe tiene como objetivo transformar información real de la universidad en un dashboard interactivo y profesional. A partir de la recolección de datos, buscamos identificar patrones para obtener insights que permitan mejorar la experiencia estudiantil y apoyar a tomar decisiones académicas. el análisis desarrollado utilizando python, con las bibliotecas pandas y dash, en el entorno de anaconda navigator.

El alcance del análisis se limita a un total de 200 datos recolectados, lo que puede representar una limitación en la representativa de los hallazgos. Sin embargo, se pretende que el dashboard resulte en una herramienta visual intuitiva que facilite la interpretación de la información y que pueda ser desplegada en la web para su acceso universal.

2. Metodología

Para alcanzar los objetivos planteados, se siguió el siguiente proceso metodológico:

- **Fuentes de datos:**
Se recolectaron datos reales mediante una encuesta sobre el sitio preferido para estudiar, almacenándolos en un archivo CSV.
- **Herramientas y técnicas empleadas:**
 - **Entorno de desarrollo:** Se utilizó anaconda navigator con python 3.10+.
 - **Limpieza y análisis de datos:** Con pandas en jupyter notebook, se realizó la carga de datos, eliminamos valores nulos, corregimos errores y detectamos outliers. Además, se calcularon medias estadísticas como la media, mediana y la moda.
 - **Vistas minables:** Se aplicaron técnicas de agrupación, filtrando resúmenes para extraer insights relevantes de la información.
 - **Desarrollo del dashboard:** Con dash y plotly, se construyó un dashboard interactivo, que integra varios gráficos.
- **Proceso de despliegue:** Finalmente, se subió el archivo a github y se desplegó en render.com permitiendo que el dashboard sea accesible desde cualquier navegador sin necesidad de tener python instalado localmente.

3. Análisis Exploratorio de Datos (EDA)

Visualizaciones y estadísticas descriptivas

- Cargamos el dataset y eliminamos los registros duplicados para garantizar la calidad de los datos.
- Se calculó la moda de la variable lugar de estudio, identificando el sitio preferido por los estudiantes. Además, se obtuvo el conteo de frecuencia para cada lugar, lo que se visualizó mediante gráfico de barras que muestra la distribución de los sitios de estudio.
- Se analizaron las horas de estudio mediante el cálculo de la media, mediana y la desviación estándar, además, se agruparon los datos por lugar de estudio y carrera para conocer el promedio de horas de estudio en cada categoría.
- Se evaluó la puntuación de los lugares de estudio, calculando el promedio general y por cada sitio. Las distribuciones de la variable puntuación se visualizaron

Detección de valores atípicos y datos faltantes

- Se aplicó la técnica del rango intercuartílico (IQR) a la variable horas de estudio para identificar valores atípicos. Se definieron los límites interiores y superiores y se filtraron los registros fuera de esos rangos.
- Se utilizaron boxplot para visualizar de forma gráfica la presencia de outliers en horas de estudio y puntuación.
- Durante este proceso de limpieza se verificó la ausencia de valores nulos y se corrigieron posibles errores de datos.

Relación entre variables relevantes

- Se exploró la relación entre carrera y lugar de estudio, identificando mediante agrupaciones el sitio de estudio más preferido para cada carrera.
- Asimismo, se analizaron las variables nuevos espacios y mejoras sugeridas:
 - Se calculó la moda de cada variable para conocer la percepción general sobre la necesidad de nuevos espacios de estudio y las mejoras sugeridas.
 - Se creó un cruce de variables y se visualizó mediante un heatmap para evidenciar cómo se relacionan las mejoras sugeridas con los diferentes sitios de estudio.
- Finalmente, se detectó un sesgo de selección en la data: la carrera de ingeniería de sistemas se sobre presenta en comparación con otras carreras, lo que podría limitar la generalización de los hallazgos al conjunto de población estudiantil.

4. Resultados y Hallazgos

- Preferencias y horas de estudio: se identificó el lugar de estudio preferido y se observó que ciertas carreras dedican más horas al estudio.
- Evaluación de puntuación: los lugares de estudio presentan, en general, buenas calificaciones, destacándose aquellos con mejores evaluaciones.
- Sugerencias de mejora: los estudiantes indican la necesidad de nuevos espacios y mejoras en los existentes, lo que se evidencia tanto en las estadísticas como en las visualizaciones.
- Sesgo de muestra: la sobre representación de estudiantes de ingeniería de sistemas sugiere que los hallazgos podrían no ser totalmente generalizables.

5. Conclusiones y Recomendaciones

- **Conclusiones:**
 - Existen diferencias notables en la elección de espacios y en las horas de estudio según la carrera.
 - Los lugares con mejores puntuaciones y aquellos que reciben sugerencias de mejora pueden orientar futuras inversiones en infraestructura.
- **Recomendaciones:**
 - Implementar mejoras en los espacios de estudio menos utilizados.
 - Ampliar la muestra para obtener una visión más representativa de todas las carreras.
 - Realizar análisis segmentados y desplegar el dashboard en línea para facilitar la toma de decisiones en tiempo real.

6. Anexos y Referencias

- Se incluye el código utilizado (Jupyter Notebook, app.py) y el archivo de datos (datos.csv).

- Se adjuntan capturas de pantalla, un video de presentación y se listan las fuentes consultadas, siguiendo la norma APA versión 7.