Exploración de Datos en la Universidad: Lugares de Estudio

Juan Sebastián Madera Coronel

Elio Fuentes Julio

Nain Faruk Chalabe Cifuentes

Universidad del Sinú

Seccional-Cartagena

Electiva Disciplinar I

Ing Cristian Cuadrado Beltrán

28 de Marzo de 2025

Resumen

El presente informe detalla el proceso de recolección, limpieza, análisis y visualización de datos acerca de los **lugares de estudio** en la universidad, utilizando Python, Pandas y Dash. Se recolectaron 158 registros reales y 142 datos ficticios sobre tiempos de espera, disponibilidad de espacios y satisfacción de los usuarios en diferentes áreas de estudio (bibliotecas, salas de estudio, cafeterías y áreas al aire libre). A partir del análisis exploratorio se identificaron patrones que sugieren que los espacios con ambiente tranquilo y buena iluminación favorecen la concentración. Finalmente, se desarrolló un dashboard interactivo que permite filtrar y comparar la eficacia de cada espacio.

Palabras clave: lugar de estudio, análisis de datos, dashboard, Python, visualización

Introducción

El estudio de los espacios de aprendizaje es fundamental para optimizar el ambiente académico y mejorar el rendimiento estudiantil. En esta actividad se exploró el tema de los **lugares de estudio** en la Universidad del Sinú, Seccional Cartagena, buscando identificar cuáles áreas facilitan una mayor concentración y productividad. La hipótesis plantea que los espacios con mayor tranquilidad, menor ruido ambiental y mejor iluminación natural son percibidos como más adecuados para estudiar.

Método

Se diseñó un estudio cuantitativo en el que se recolectaron datos reales de 158 registros de lugares de estudio en la Universidad. Las variables incluyeron:

- **Nombre:** Identificación personal (Nombre y apellido).
- **Carrera:** Varían entre Gastronomía, Derecho, Odontología, Ing Industrial, Medicina, Ing Sistemas, Ing Civil, Fisioterapia, Psicología, entre otras.
- Jornada: Entre la jornada Diurna y Nocturna.
- **Semestre:** Entre Primer semestre hasta Decimo semestre.
- Edad: Entre 16 años hasta los 30 años.
- Lugar: Cafetería, Biblioteca, Salas informáticas, Salones, Pasillo y Sala de redes.
- **Género:** Masculino y Femenino.

Los datos fueron cargados en un archivo CSV y posteriormente importados a Python usando la biblioteca Pandas. Se procedió a limpiar el dataset eliminando valores nulos y corrigiendo errores de ingreso. Se calcularon medidas descriptivas (media, mediana, moda y desviación estándar) y se detectaron outliers mediante boxplots.

Resultado

El análisis revelo que las Bibliotecas y las Cafeterías presentan más demanda para el uso de estás como lugar de estudio con mayor satisfacción para los Estudiantes, en comparación con los salones o salas de Redes. Un dashboard interactivo, desarrollado con Dash, permite visualizar la distribución de preferencias de lugares, valores Atípicos, Diagramas de Dispersión y el Diagrama de Frecuencia de lugares. Al momento de la recolección de datos pudimos notar que estos lugares eran escogidos especialmente por la tranquilidad, menor ruido ambiental y mejor iluminación que le brinda a los Estudiantes.

Discusión

Los hallazgos apoyan la hipótesis inicial: los espacios de estudio tranquilos, silenciosos y bien iluminados son preferidos por los estudiantes. Estos resultados son consistentes con estudios previos (Juan Barón, 2024; Juan Aidar 2024) Y subrayan la importancia de diseñar ambientes que promuevan el aprendizaje. Se recomienda que la Universidad considere estos factores al planificar mejoras en la infraestructura. Además el dashboard desarrollado permite una visualización dinámica de los datos, facilitando la toma de decisiones basadas en evidencias.

Conclusión

El presente estudio permitió identificar y analizar las características de los lugares de estudio en la Universidad de Ejemplo. La integración de técnicas de análisis exploratorio y visualización mediante Python y Dash ofreció una herramienta interactiva para evaluar y comparar diferentes espacios, lo que puede ser de utilidad para futuras estrategias de optimización en ambientes académicos.

Apéndices

Apéndice A: Captura de pantalla del Dashboard

Análisis de Preferencias de Lugares

Lugar más frecuente: Biblioteca

► Total de registros en 'Lugar': 300

► Media de frecuencia de lugares: 54.61 - 55

► Mediana de frecuencia de lugares: 55.00 - Salas informaticas

Desviación estándar: 13.89

► Q1 (Primer cuartil): 47.00

► Q3 (Tercer cuartil): 59.00

► IQR (Rango Intercuartil): 12.00

► Límite Inferior: 29.00

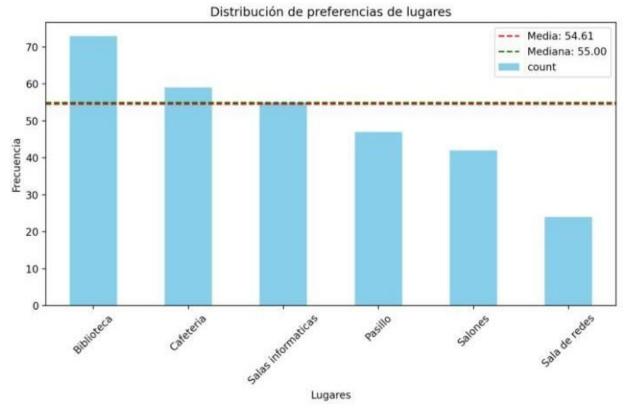
► Límite Superior: 77.00

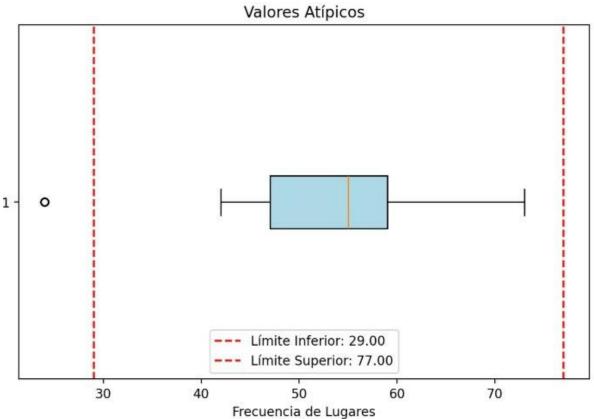
★ Valores Atípicos

```
Sala de redes
```

Distribución de Frecuencias

```
Lugar
Biblioteca 73
Cafeteria 59
Salas informaticas 55
Pasillo 47
Salones 42
Sala de redes 24
```

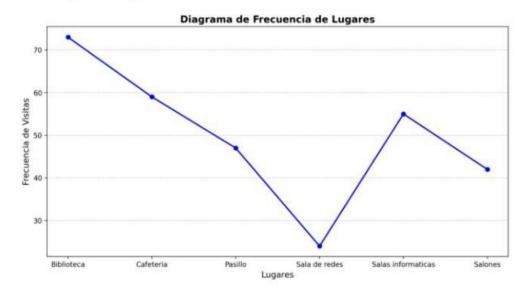


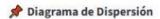


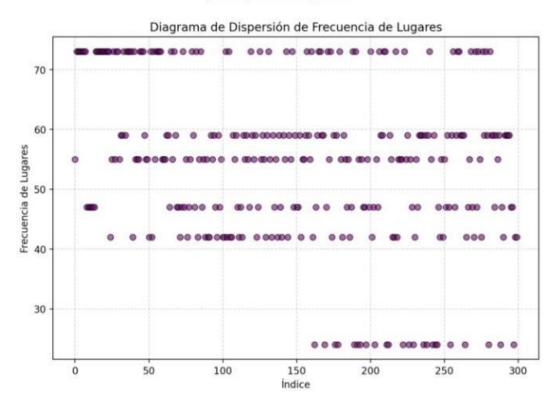
Analisis de Preferencias de Lugares

Lugar más frecuente: Biblioteca

Total de registros en 'Lugar': 300







Apéndice B: Código Fuente (Fragmentos Relevantes)

```
st.code(frecuencia_lugares.to_string())
                                                                                              Aa ab * ? of 12
                                                                                                                  \wedge \downarrow = \times
                                                                     > fig
# Gráfico de barras
fig, ax = plt.subplots(figsize=(10, 5))
frecuencia_lugares.plot(kind='bar', color='skyblue', ax=ax)
ax.axhline(media_valor, color='red', linestyle='dashed', label=f'Media: {media_valor:.2f}')
ax.axhline(mediana_valor, color='green', linestyle='dashed', label=f'Mediana: {mediana_valor:.2f}')
ax.set_title("Distribución de preferencias de lugares")
ax.set_xlabel("Lugares")
ax.set_ylabel("Frecuencia")
ax.legend()
ax.set_xticklabels(ax.get_xticklabels(), rotation=45)
st.pyplot(fig)
fig2, ax2 = plt.subplots(figsize=(8, 5))
ax2.boxplot(df["Lugar_numeric"], vert=False, patch_artist=True, boxprops=dict(facecolor="lightblue"))
ax2.axvline(limite_inferior, color="red", linestyle="dashed", label=f"Límite_Inferior: {limite_inferior:.2f}")
ax2.axvline(limite_superior, color="red", linestyle="dashed", label=f"Límite Superior: {limite_superior:.2f}")
ax2.legend()
ax2.set_title("Valores Atípicos")
ax2.set_xlabel("Frecuencia de Lugares")
st.pyplot(fig2)
st.markdown('★ Diagrama de Dispersión', unsafe allow html=True)
fig3, ax3 = plt.subplots(figsize=(9, 6))
ax3.scatter(df.index, df["Lugar_numeric"], color='purple', alpha=0.6, edgecolors='black')
ax3.set_title("Diagrama de Dispersión de Frecuencia de Lugares")
ax3.set_xlabel("Índice")
ax3.set_ylabel("Frecuencia de Lugares")
ax3.grid(True. linestvle="--". alpha=0.5)
                                                                 Ln 6, Col 32 Spaces: 4 UTF-8 CRLF () Python 🔠 3.10.16 ('streamlit': cor
                                                                 > fig
                                                                                      Aa <u>ab</u> * ? of 12
                                                                                                        \wedge \downarrow = \times
    df = pd.read_excel(archivo_excel)
    df["Lugar"] = df["Lugar"].fillna('').astype(str).str.strip().str.capitalize()
     frecuencia_lugares = df["Lugar"].value_counts()
    total_registros = df["Lugar"].count()
    df["Lugar_numeric"] = df["Lugar"].map(frecuencia_lugares)
    media_valor = df["Lugar_numeric"].mean()
    mediana_valor = df["Lugar_numeric"].median()
    desviacion_std = df["Lugar_numeric"].std()
    moda = df["Lugar"].mode()[0]
     mediana_lugar = frecuencia_lugares[frecuencia_lugares == mediana_valor].index[0] if (frecuencia_lugares == mediana_valor
     media_lugar = frecuencia_lugares.iloc[(frecuencia_lugares - media_valor).abs().argsort()[0]]
    Q1, Q3 = df["Lugar_numeric"].quantile([0.25, 0.75])
    IOR = 03 - 01
    limite_inferior = max(0, Q1 - 1.5 * IQR)
limite_superior = Q3 + 1.5 * IQR
     valores_atipicos = df["Lugar"][(df["Lugar_numeric"] < limite_inferior) | (df["Lugar_numeric"] > limite_superior)]
    st.markdown('div class="box centered")', unsafe_allow_html=True)
st.markdown(f"** Lugar más frecuente:** {moda}")
st.markdown(f"** Total de registros en 'Lugar':** {total_registros}")
     st.markdown(f"*** Media de frecuencia de lugares:** {media_valor:.2f} - {media_lugar}")
     st.markdown(f"**▶ Mediana de frecuencia de lugares:** {mediana_valor:.2f} - {mediana_lugar}")
     st.markdown(f"** Desviación estándar:** {desviacion_std:.2f}")
```