Gies Business and Online Customer Action Data Analysis

Load Required Libraries and Data Extracted from Google Analytics

```
1   import pandas as pd
2
3   # Attempt to read with ISO-8859-1 encoding
4   business_df = pd.read_csv("Gies_Business_Customer_Journey.csv",
    encoding="ISO-8859-1")
5   online_df = pd.read_csv("Gies_Online_Customer_Journey.csv",
    encoding="ISO-8859-1")
6
7   # Display the first few rows
8   print(business_df.head())
9   print(online_df.head())
10
11  business_df = business_df.dropna()
12  online_df = online_df.dropna()
13
14  # Check for missing values
15  business_df.isnull().sum()
16  online_df.isnull().sum()
```

```
       Date Date Final        Event name  \
0  20250101  01-Jan-25         page_view
1  20250101  01-Jan-25  user_engagement
2  20250101  01-Jan-25    session_start
3  20250101  01-Jan-25  user_engagement
4  20250101  01-Jan-25         page_view

                                Page location  \
0              https://giesbusiness.illinois.edu/
1              https://giesbusiness.illinois.edu/
2              https://giesbusiness.illinois.edu/
3  https://giesbusiness.illinois.edu/faculty-rese...
4              https://giesbusiness.illinois.edu/msf

                      Page title and screen name  Event count  Total users
0  Gies College of Business | University of Illin...          231          189
1  Gies College of Business | University of Illin...          199          152
2  Gies College of Business | University of Illin...          153          147
3  Faculty Profiles | Gies College of Business | ...          112           31
4  Masters in Finance | Gies College MSF Program ...          109           80
       Date Event name                              Page title  \
0  20250113      click  ACCY 501: Accounting Analysis I | Gies Online ...
1  20250113      click  ACCY 503: Managerial Accounting | Gies Online ...
2  20250113      click       ACCY 504: Auditing | Gies Online | Illinois
3  20250113      click  ACCY 506: Advanced Topics in Accounting | Gies...
4  20250113      click  ACCY 517: Financial Statement Analysis and Val...

                                Page location  Event count  Total users
0  https://giesonline.illinois.edu/courses/accy-5...            1            1
1  https://giesonline.illinois.edu/courses/accy-5...            6            3
2  https://giesonline.illinois.edu/courses/accy-5...            3            2
3  https://giesonline.illinois.edu/courses/accy-5...            1            1
4  https://giesonline.illinois.edu/courses/accy-5...            2            1
```

	0
Date	0
Event name	0
Page title	0
Page location	0
Event count	0
Total users	0

dtype: int64

**Top 10 Events in Business vs. Online Journeys**

```python
1  # For the business dataset
2  business_event_counts = business_df['Event name'].value_counts().reset_index()
3  business_event_counts.columns = ['Event name', 'Frequency']
4  print(business_event_counts.head(10))  # Top 10 events
5
6  # For the online dataset
7  online_event_counts = online_df['Event name'].value_counts().reset_index()
8  online_event_counts.columns = ['Event name', 'Frequency']
9  print(online_event_counts.head(10))    # Top 10 events
10
11 import matplotlib.pyplot as plt
12 import seaborn as sns
13
14 # Plot top 10 events for business
15 top_10_business = business_event_counts.head(10)
16 plt.figure(figsize=(10,6))
17 sns.barplot(data=top_10_business, x='Frequency', y='Event name', color='blue')
18 plt.title('Top 10 Events (Business)')
19 plt.xlabel('Frequency')
20 plt.ylabel('Event Name')
21 plt.show()
22
23 # Plot top 10 events for online
24 top_10_online = online_event_counts.head(10)
25 plt.figure(figsize=(10,6))
26 sns.barplot(data=top_10_online, x='Frequency', y='Event name', color='orange')
27 plt.title('Top 10 Events (Online)')
28 plt.xlabel('Frequency')
29 plt.ylabel('Event Name')
30 plt.show()
31
```

```
      Event name  Frequency
0          page_view     201686
1      session_start     173895
2        first_visit     133023
3    user_engagement      80036
4             scroll      24069
5              click       8723
6  view_search_results    2241
7      file_download       1884
8    click_apply_grad       629
9       generate_lead       393
      Event name  Frequency
0          page_view     247204
1      session_start     207333
2        first_visit     158973
3    user_engagement      48830
4             scroll      26995
5              click       4715
6         form_start        986
7        video-start        766
8        form_submit        680
9       generate_lead       680
```

Top 10 Events (Business)

Top 10 Events (Online)
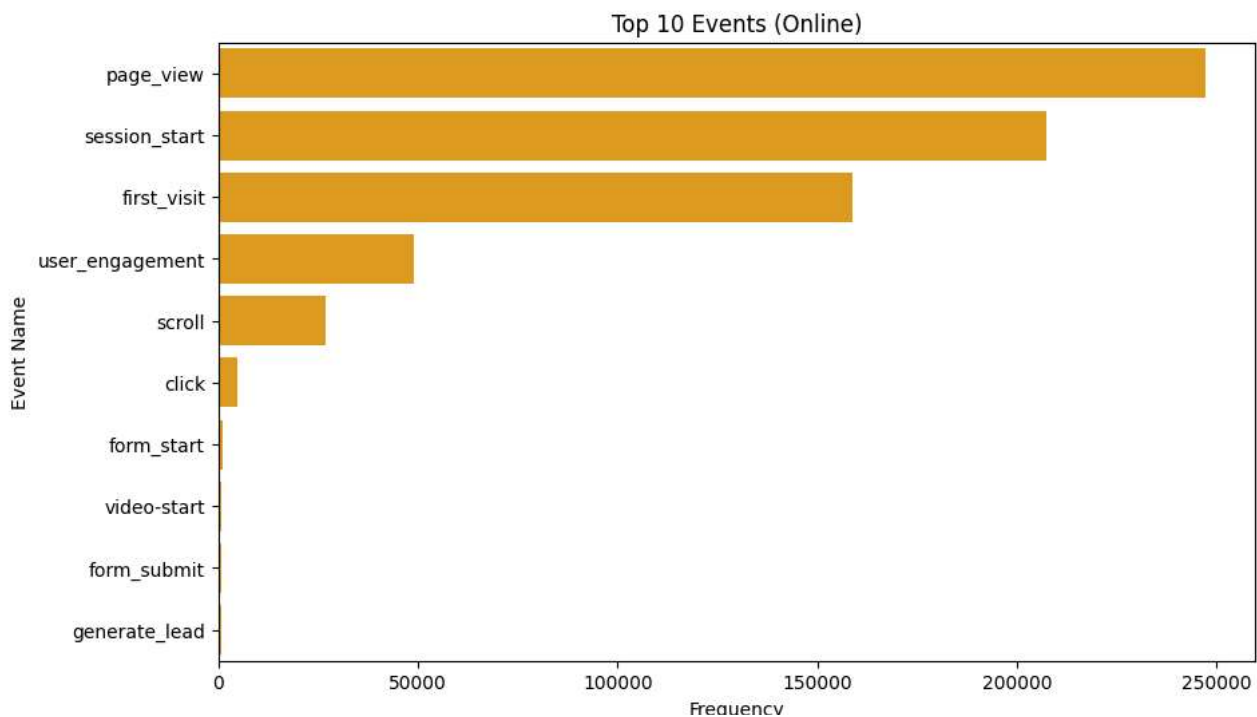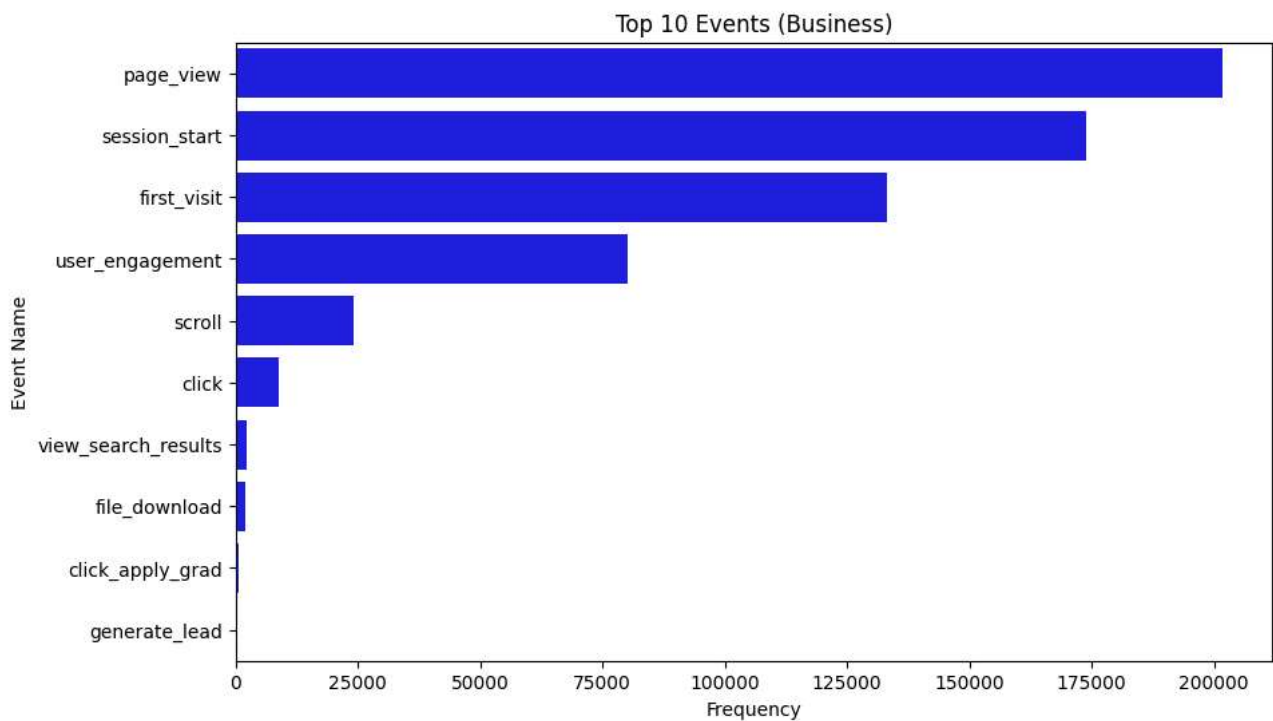
ANALYSIS

"page_view," "session_start," and "first_visit" dominate both datasets, showing a focus on initial visits and page views. In Business, events like "file_download" and "click_apply_grad" appear, while Online features "form_start" and "video-start," reflecting different user pathways. "generate_lead" appears in both but less frequently, suggesting a smaller subset of users proceed to lead-generating actions.

Analyzing Top Page Locations

```python
1 # For the business dataset
2 business_page_counts = (
3     business_df.groupby('Page location', as_index=False)['Event count'].sum()
4     .sort_values(by='Event count', ascending=False)
5 )
6
7 # For the online dataset
8 online_page_counts = (
9     online_df.groupby('Page location', as_index=False)['Event count'].sum()
10     .sort_values(by='Event count', ascending=False)
11 )
12
13 # Display the top rows
14 print("Top Business Page Locations:")
15 print(business_page_counts.head(10))
16
17 print("\nTop Online Page Locations:")
18 print(online_page_counts.head(10))
19
20 import matplotlib.pyplot as plt
21 import seaborn as sns
22
23 # Take the top 10 for Business
24 top_10_business_pages = business_page_counts.head(10)
25 plt.figure(figsize=(10,6))
26 sns.barplot(data=top_10_business_pages, x='Event count', y='Page location', color='green')
27 plt.title('Top 10 Page Locations (Business)')
28 plt.xlabel('Total Event Count')
29 plt.ylabel('Page Location')
30 plt.show()
31
32 # Take the top 10 for Online
33 top_10_online_pages = online_page_counts.head(10)
34 plt.figure(figsize=(10,6))
35 sns.barplot(data=top_10_online_pages, x='Event count', y='Page location', color='red')
36 plt.title('Top 10 Page Locations (Online)')
37 plt.xlabel('Total Event Count')
38 plt.ylabel('Page Location')
39 plt.show()
40
```
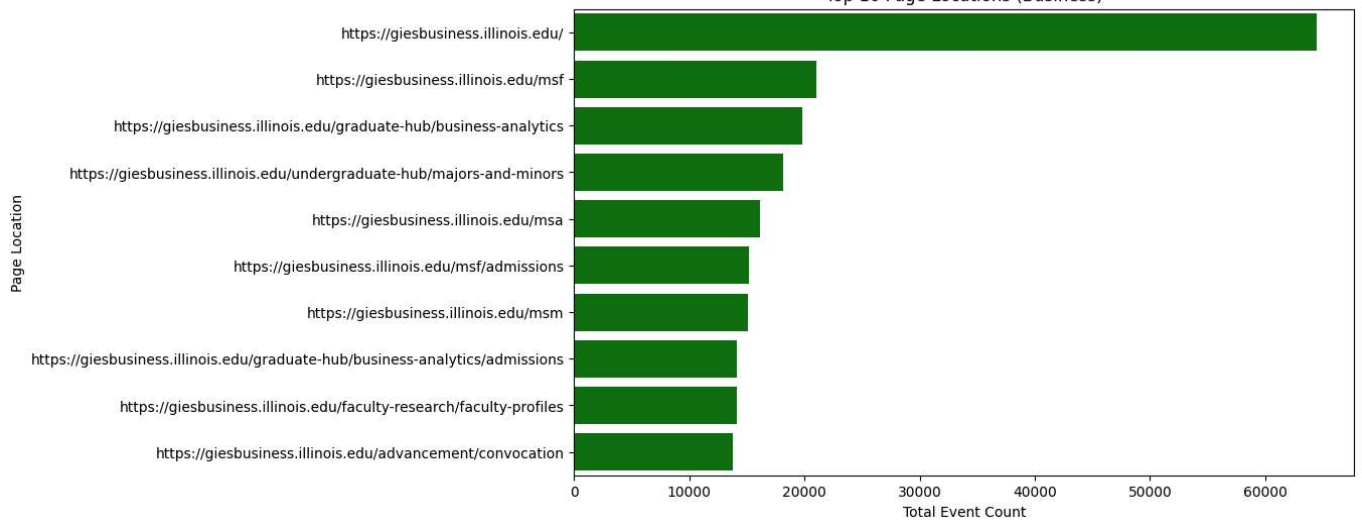
⇥ Top Business Page Locations:

|	Page location	Event count
42	https://giesbusiness.illinois.edu/	64494
91358	https://giesbusiness.illinois.edu/msf	21005
19534	https://giesbusiness.illinois.edu/graduate-hub...	19845
131750	https://giesbusiness.illinois.edu/undergraduat...	18118
82526	https://giesbusiness.illinois.edu/msa	16161
91360	https://giesbusiness.illinois.edu/msf/admissions	15172
102339	https://giesbusiness.illinois.edu/msm	15108
19535	https://giesbusiness.illinois.edu/graduate-hub...	14139
19431	https://giesbusiness.illinois.edu/faculty-rese...	14135
18868	https://giesbusiness.illinois.edu/advancement/...	13750

Top Online Page Locations:

	Page location	Event count
153001	https://giesonline.illinois.edu/explore-progra...	71351
153194	https://giesonline.illinois.edu/explore-progra...	29216
153003	https://giesonline.illinois.edu/explore-progra...	29122
127356	https://giesonline.illinois.edu/explore-progra...	19530
47	https://giesonline.illinois.edu/	18020
106044	https://giesonline.illinois.edu/explore-progra...	16391
88429	https://giesonline.illinois.edu/explore-progra...	16037
218239	https://giesonline.illinois.edu/explore-progra...	14014
220469	https://giesonline.illinois.edu/tuition-costs	11071
127462	https://giesonline.illinois.edu/explore-progra...	10860

Top 10 Page Locations (Business)

Top 10 Page Locations (Online)
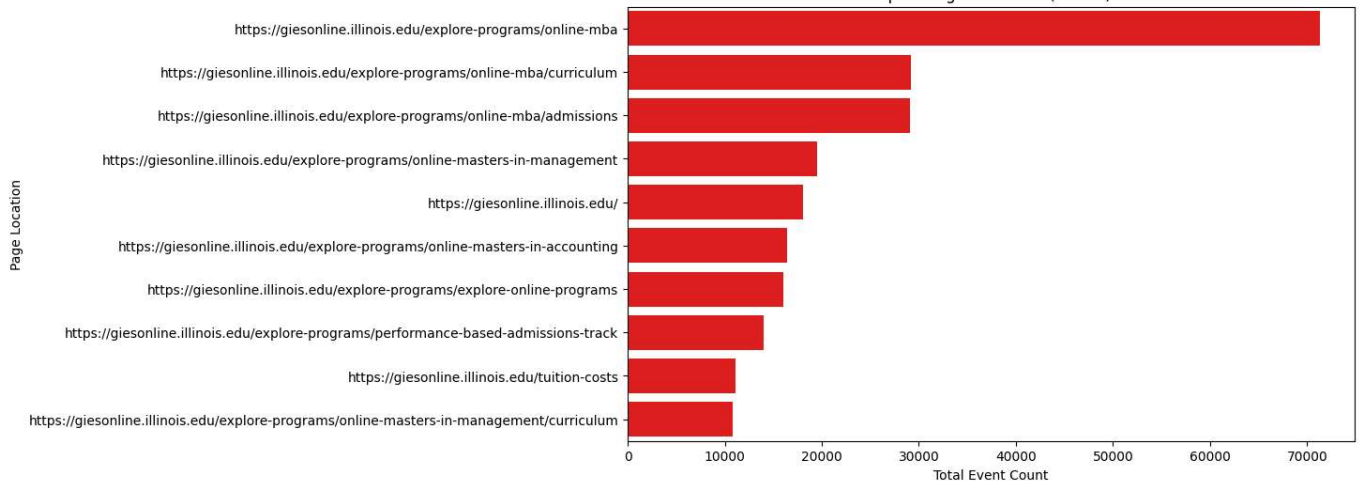
**Top Page Locations: Concise Observations**

- **Gies Business**:

  The main homepage draws the most traffic, serving as a central hub. Specialized graduate program pages (e.g., MSF, Business Analytics, MSBA) also see high engagement, indicating strong interest in advanced studies. Admissions and undergrad "majors-and-minors" pages rank well, reflecting interest across multiple academic levels.

- **Gies Online**:

  The iMBA and its curriculum pages dominate, highlighting the program's appeal. Admissions, performance-based admissions tracks, and tuition cost pages also rank high, emphasizing prospective students' focus on enrollment feasibility and affordability.

Comparing User Engagement by Event Type Over Time

```python
1 # Example events to analyze
2 events_of_interest = ["page_view", "session_start", "form_submit"]
3 # Filter Business dataset for selected events
4 business_filtered = business_df[business_df['Event name'].isin(events_of_interest)]
5
6 # Group by date and event name, then sum event counts
7 business_event_trends = (
8     business_filtered
9     .groupby(['Date', 'Event name'], as_index=False)['Event count']
10     .sum()
11 )
12
13 # Same process for Online dataset
14 online_filtered = online_df[online_df['Event name'].isin(events_of_interest)]
15 online_event_trends = (
16     online_filtered
17     .groupby(['Date', 'Event name'], as_index=False)['Event count']
18     .sum()
19 )
20 import seaborn as sns
21 import matplotlib.pyplot as plt
22
23 plt.figure(figsize=(12,6))
24 sns.lineplot(data=business_event_trends, x='Date', y='Event count', hue='Event name', palette='tab10')
25 plt.title("Business: Key Events Over Time")
26 plt.xlabel("Date")
27 plt.ylabel("Total Event Count")
28 plt.xticks(rotation=45)
29 plt.legend(title="Event Name")
30 plt.show()
31
32 plt.figure(figsize=(12,6))
33 sns.lineplot(data=online_event_trends, x='Date', y='Event count', hue='Event name', palette='tab10')
34 plt.title("Online: Key Events Over Time")
35 plt.xlabel("Date")
36 plt.ylabel("Total Event Count")
37 plt.xticks(rotation=45)
38 plt.legend(title="Event Name")
39 plt.show()
40
```

Business: Key Events Over Time

Online: Key Events Over Time
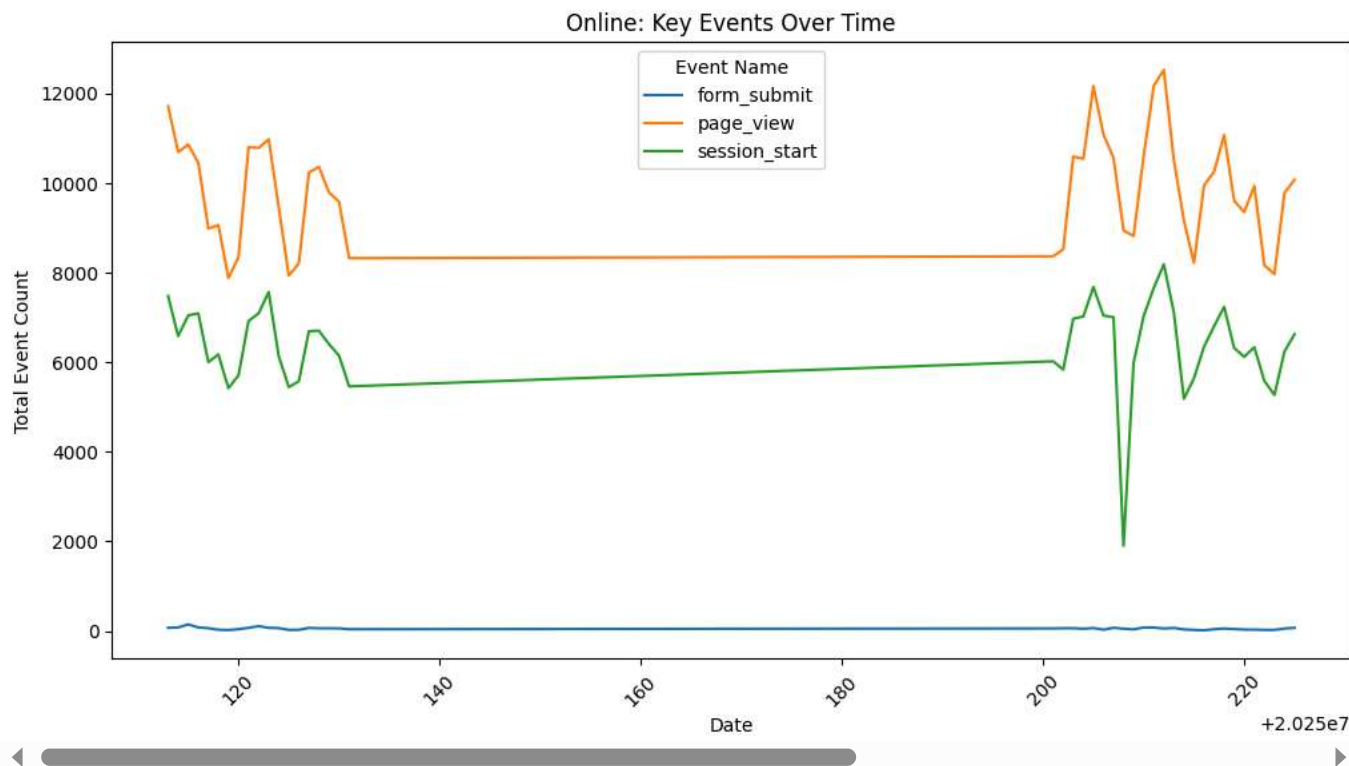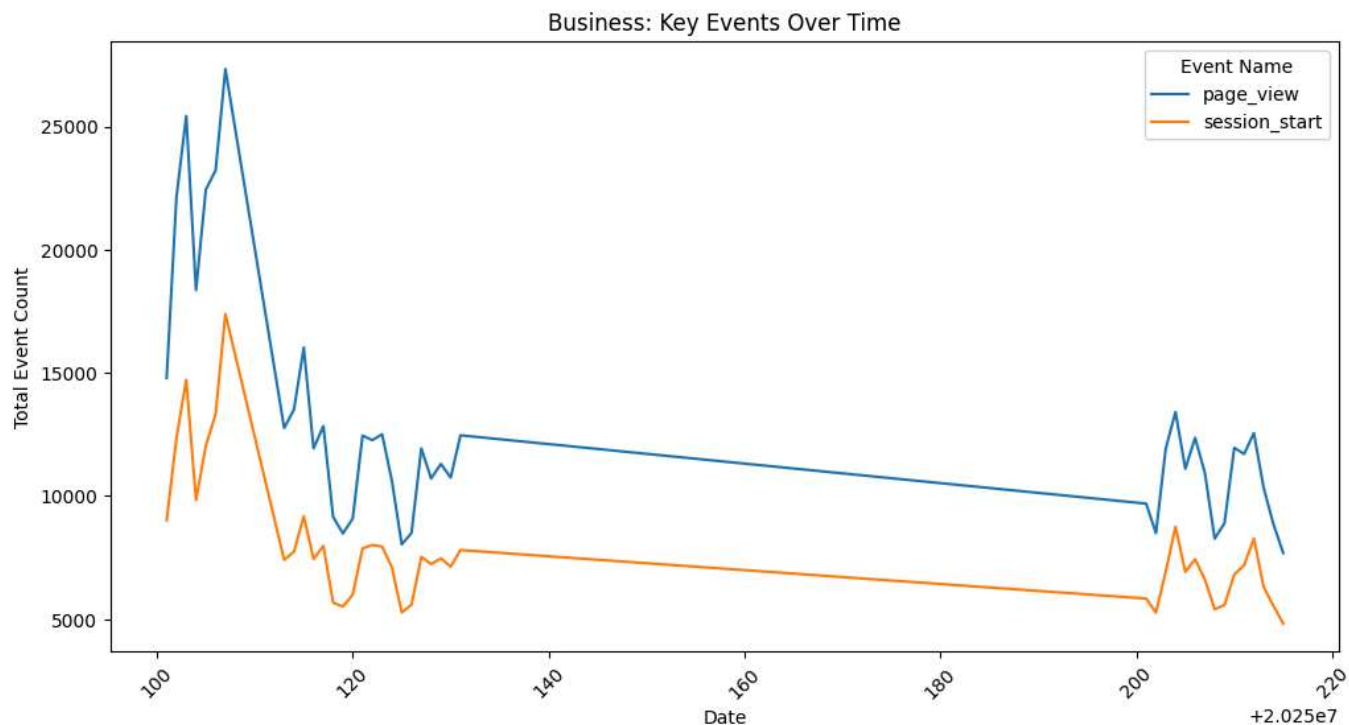
Business Dataset

"page_view" consistently outperforms "session_start", indicating users typically browse multiple pages once a session begins. Both metrics show a gradual decline over time, with occasional spikes suggesting possible promotional or academic calendar influences.

Online Dataset

"page_view" again dominates, while "form_submit" remains minimal, pointing to a lower direct conversion rate or potential data-collection gaps. "session_start" trends remain below "page_view" but show moderate fluctuation, hinting at consistent user inflow with varying engagement levels.

High-Level Machine Learning Approach: User Segmentation

```
1   import pandas as pd
```