kaggle.com/code/geetanjali11kumari11/merchalytics-google-store-insights/edit

All Bookmarks

Run All Markdown Draft Session off (run a cell to start)

Checking the Database

First of all, let's have a look at the database and see how the information is stored.

+ Code + Markdown

[4]:

```
import numpy as np
import pandas as pd
from google.cloud import bigquery
import bq_helper
from bq_helper import BigQueryHelper
import matplotlib.pyplot as plt
import seaborn as sns

google_analytics = bq_helper.BigQueryHelper(active_project="bigquery-public-data",
                                dataset_name="data:google_analytics_sample")

bq_assistant = BigQueryHelper("bigquery-public-data", "google_analytics_sample")
bq_assistant.head("ga_sessions_20170101", num_rows=3)
```

```
Using Kaggle's public dataset BigQuery integration.
Using Kaggle's public dataset BigQuery integration.
```

[4]:

	visitorId	visitNumber	visitId	visitStartTime	date	totals	trafficSource	device	geoNetwork	customDimensions	hits	fullVisitorId	userId	channelGrouping	socialEngagementType
0	None	2	1483290878	1483290878	20170101	{'visits': 1, 'hits': 2, 'pageviews': 1, 'time...	{'referralPath': None, 'campaign': '(not set)'...	{'browser': 'Chrome', 'browserVersion': 'not a...	{'continent': 'Europe', 'subContinent': 'South...	[{'index': 4, 'value': 'EMEA'}]	[{'hitNumber': 1, 'time': 0, 'hour': 9, 'minut...	7431279462169656568	None	Organic Search	Not Socially Engaged
1	None	1	1483293597	1483293597	20170101	{'visits': 1, 'hits': 2, 'pageviews': 2, 'time...	{'referralPath': '/YKEI_mrn/items/c10b14f9a69f...	{'browser': 'Safari', 'browserVersion': 'not a...	{'continent': 'Asia', 'subContinent': 'Eastern...	[{'index': 4, 'value': 'APAC'}]	[{'hitNumber': 1, 'time': 0, 'hour': 9, 'minut...	1336484329946561874	None	Referral	Not Socially Engaged
2	None	1	1483292307	1483292307	20170101	{'visits': 1, 'hits': 2, 'pageviews': 2, 'time...	{'referralPath': None, 'campaign': '(not set)'...	{'browser': 'Chrome', 'browserVersion': 'not a...	{'continent': 'Americas', 'subContinent': 'Nor...	[{'index': 4, 'value': 'North America'}]	[{'hitNumber': 1, 'time': 0, 'hour': 9, 'minut...	1701623065972643878	None	Organic Search	Not Socially Engaged

Main Traffic Sources

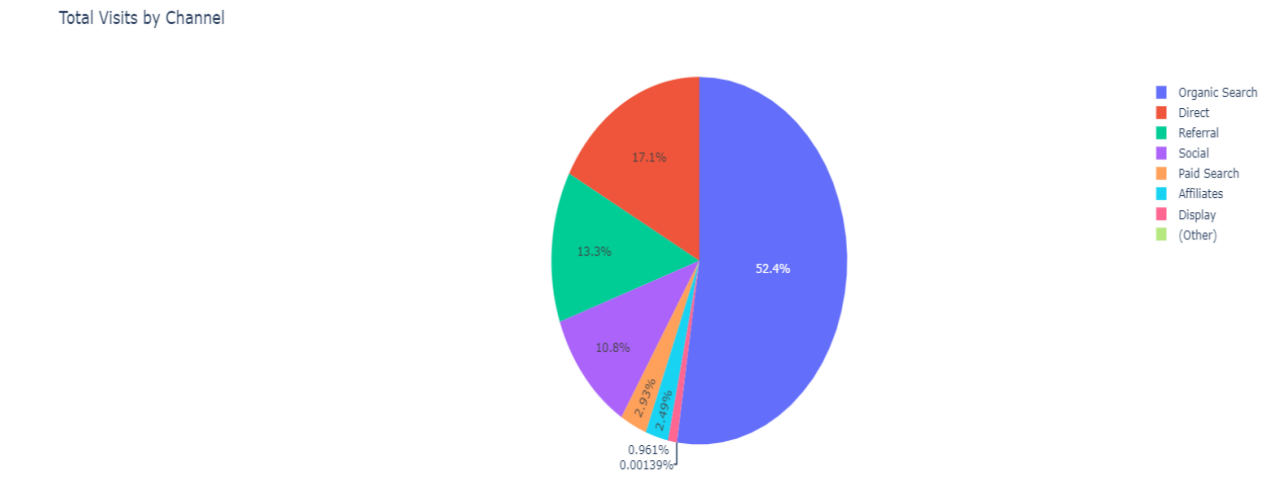
What were the main traffic sources driving visitors to the online store [in July 2017]?

By checking the main traffic sources, we can find out more about the store's approach to marketing and how well or how bad some sources are performing. There is also a month-on-month detailed information for each source later in this project, when we try to identify growth trends.

[10]:

```
import plotly.express as px
import plotly.graph_objects as go
import plotly.subplots as sp

query = """SELECT
channelGrouping,
COUNT (visitId ) AS number_of_visits,
FROM `bigquery-public-data.google_analytics_sample.ga_sessions_*`
WHERE
_TABLE_SUFFIX BETWEEN '20170701' AND '20170731'
GROUP BY
channelGrouping
HAVING number_of_visits > 0
ORDER BY
number_of_visits DESC;
        """
response = google_analytics.query_to_pandas_safe(query)

px.pie(data_frame=response,
      names='channelGrouping',
      values='number_of_visits',
      title='Total Visits by Channel')
```

Total Visits by Channel



+ Code + Markdown

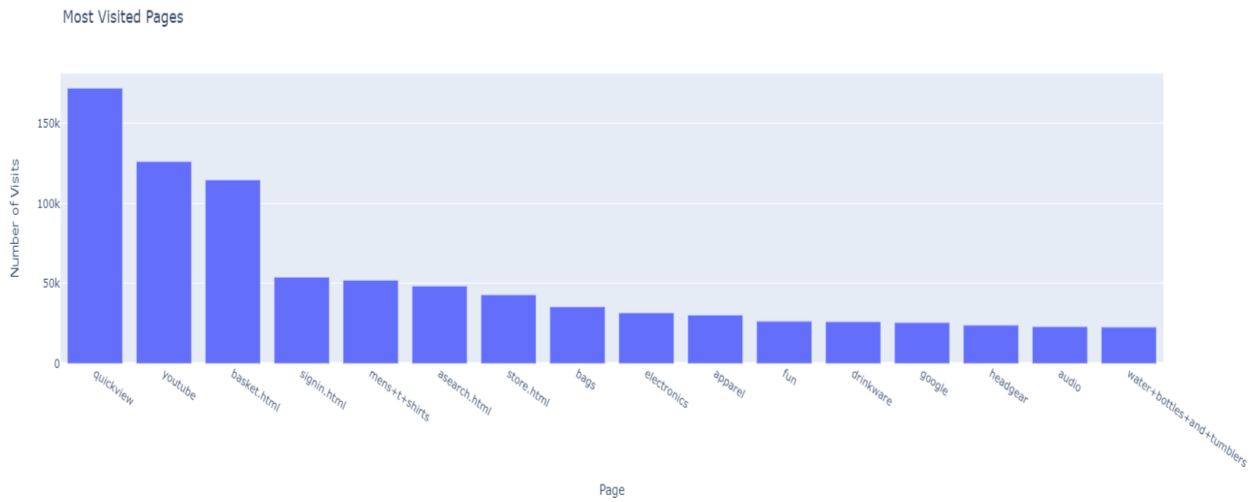
Most Visited Pages

What are the most visited pages in the online store website?

Knowing these pages can show us which products are the most sought after by online customers, and this information can help the online store define its content marketing and SEO strategy.

[11]:

```
query = """
SELECT
  SUBSTRING(h.page.pagePath, INSTR(h.page.pagePath, '/', -1) + 1) AS page,
  COUNT(h.page.pagePath) AS number_of_visits
FROM
  `bigquery-public-data.google_analytics_sample.ga_sessions_*`, UNNEST(hits) as h
WHERE
  _TABLE_SUFFIX BETWEEN '20170101' AND '20170731'
GROUP BY page
ORDER BY number_of_visits DESC;
"""

response = google_analytics.query_to_pandas_safe(query)

# Create a line plot using Plotly Express
fig = px.bar(response[1:17],
             x='page',
             y='number_of_visits',
             title='Most Visited Pages')

fig.update_xaxes(title_text='Page')
fig.update_yaxes(title_text='Number of Visits')

# Show the interactive plot
fig.show()
```



How do user demographics and behaviors impact conversion rates?

Merchalytics:Google Store Insig... Draft saved

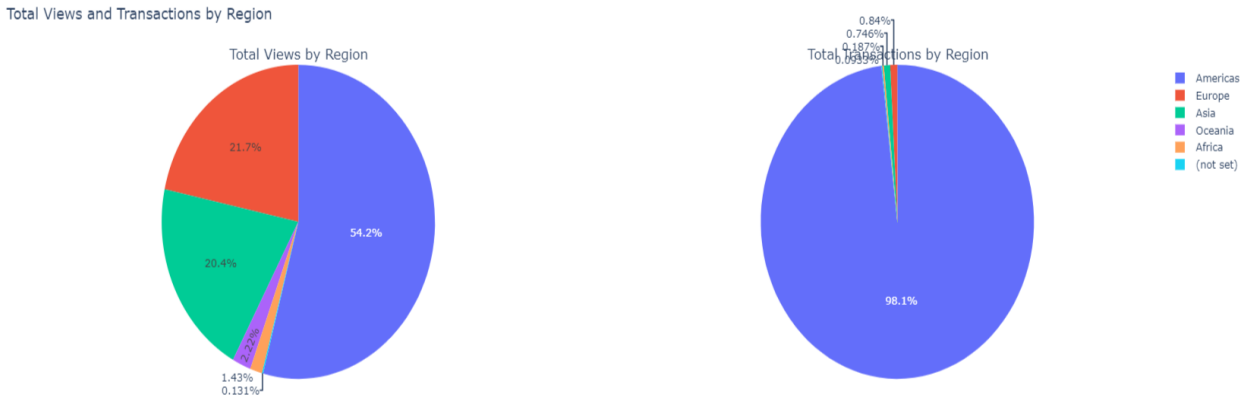
File Edit View Run Add-ons Help

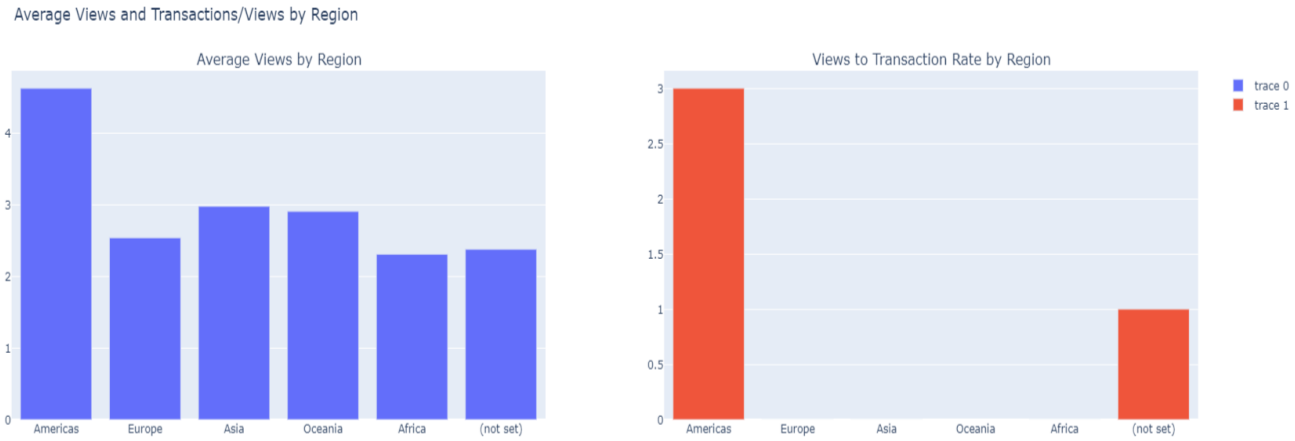
Run All Markdown Draft Session off (run a cell to start)

[12]:

```
query = """SELECT
geoNetwork.continent AS region,
SUM(totals.transactions) AS total_transactions,
COUNT(totals.pageviews) AS total_views,
COUNT(totals.bounces) AS total_bounces,
ROUND(AVG(totals.transactions), 2) AS avg_transactions,
ROUND(AVG(totals.pageviews),2) AS avg_views,
ROUND(SUM(totals.transactions) / COUNT(totals.pageviews), 2) * 100 AS views_to_transactions
FROM `bigquery-public-data.google_analytics_sample.ga_sessions_*`
WHERE
_TABLE_SUFFIX BETWEEN '20170701' AND '20170731'
GROUP BY geoNetwork.continent
ORDER BY total_views DESC;
        """

response = google_analytics.query_to_pandas_safe(query)

# Create a subplot with two pie charts
fig1 = sp.make_subplots(rows=1,
                        cols=2,
                        subplot_titles=['Total Views by Region', 'Total Transactions by Region'],
                        specs=[[{'type': 'pie'}, {'type': 'pie'}]])

# Pie chart for total views by region
fig1.add_trace(go.Pie(labels=response['region'], values=response['total_views']), row=1, col=1)

# Pie chart for total bounces by region
fig1.add_trace(go.Pie(labels=response['region'], values=response['total_transactions']), row=1, col=2)

# Update layout
fig1.update_layout(title_text="Total Views and Transactions by Region")

# Show the combined figure
fig1.show()

# Create a subplot with two pie charts
fig2 = sp.make_subplots(rows=1,
                        cols=2,
                        subplot_titles=['Average Views by Region', 'Views to Transaction Rate by Region'],
                        specs=[[{'type': 'bar'}, {'type': 'bar'}]])

# Pie chart for total views by region
fig2.add_trace(go.Bar(x=response['region'], y=response['avg_views']), row=1, col=1)
```

```
# Pie chart for total bounces by region
fig2.add_trace(go.Bar(x=response['region'], y=response['views_to_transactions']), row=1, col=2)

# Update layout
fig2.update_layout(title_text="Average Views and Transactions/Views by Region")

# Show the combined figure
fig2.show()
```





+ Code + Markdown

Daily Trends

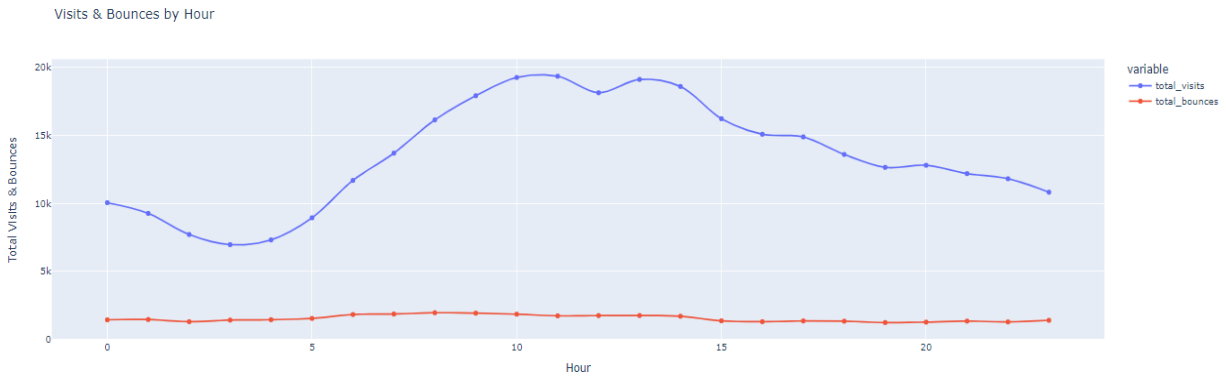
What is the daily website traffic?

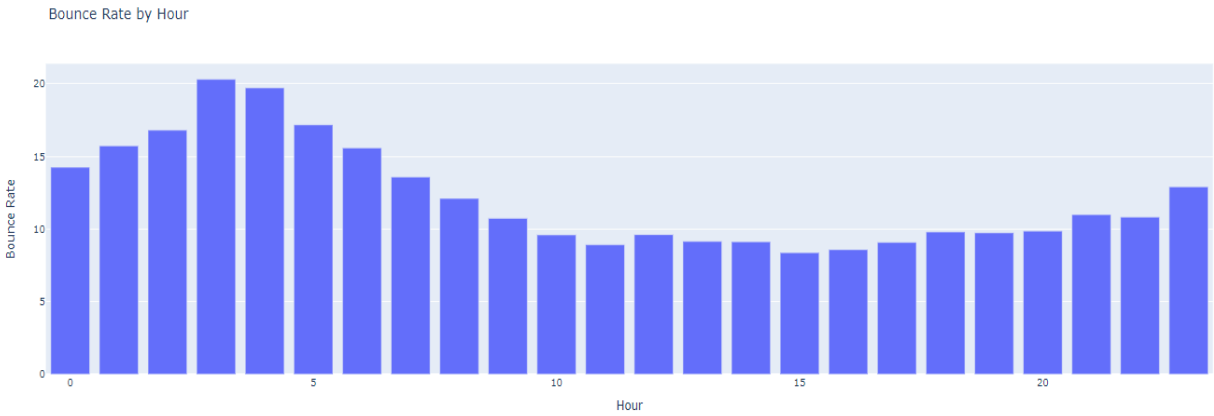
Let's have a look at the average visits and bounces. Doing this can help us understand better the user behavior and might provide some insights about the best time for creating social media posts with content that might attract more users, such as promotions and discounts.

[13]:

```
query = """
SELECT
  h.hour AS hour,
  COUNT(*) AS total_visits,
  SUM(totals.bounces) AS total_bounces,
  (SUM(totals.bounces) / COUNT(*) * 100) AS bounce_rate
FROM
  `bigquery-public-data.google_analytics_sample.ga_sessions_*`, UNNEST(hits) as h
WHERE
  _TABLE_SUFFIX BETWEEN '20170701' AND '20170731'
GROUP BY
  hour
ORDER BY
  hour;
"""

response = google_analytics.query_to_pandas_safe(query)

# Create a line plot using Plotly Express
fig = px.line(response,
              x='hour',
              y=['total_visits', 'total_bounces'],
              title='Visits & Bounces by Hour',
              markers=True,
              line_shape='spline')

fig.update_xaxes(title_text='Hour')
fig.update_yaxes(title_text='Total Visits & Bounces')

# Show the interactive plot
fig.show()

# Create a line plot using Plotly Express
fig2 = px.bar(response,
```

```
# Show the interactive plot
fig.show()

# Create a line plot using Plotly Express
fig2 = px.bar(response,
              x='hour',
              y='bounce_rate',
              title='Bounce Rate by Hour')

fig2.update_xaxes(title_text='Hour')
fig2.update_yaxes(title_text='Bounce Rate')

# Show the interactive plot
fig2.show()
```





+ Code + Markdown

Traffic Trends & Patterns

Are there any trends or patterns in website traffic and sales?

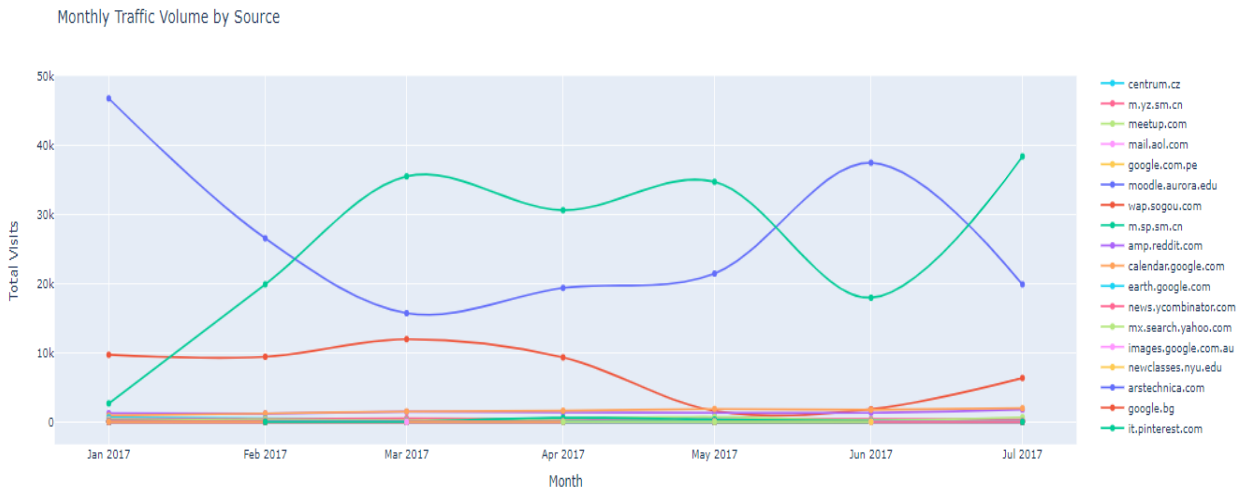
Traffic Trends & Patterns

Are there any trends or patterns in website traffic and sales?

By analyzing historical website traffic and sales data over a prolonged period, we can determine whether there are any identifiable patterns that repeat annually or at specific intervals. To analyze seasonal trends, we can aggregate data on a monthly, quarterly, or weekly basis and use visualizations such as line charts, bar graphs, or heatmaps. These visual representations can reveal patterns that are not immediately apparent when looking at individual data points. Wecan also use seasonal decomposition methods, such as moving averages or seasonal indices to extract underlying patterns from the noise.

[14]:

```
query = """
SELECT
  DATE_TRUNC(PARSE_DATE('%Y%m%d', _TABLE_SUFFIX), MONTH) AS month,
  trafficSource.source AS source,
  COUNT(*) AS total_visits
FROM
  `bigquery-public-data.google_analytics_sample.ga_sessions_*`
WHERE
  _TABLE_SUFFIX BETWEEN '20170101' AND '20170731'
GROUP BY
  month, source
ORDER BY
  month, total_visits DESC;
"""

response = google_analytics.query_to_pandas_safe(query)

# Create a line chart using Plotly Express
fig = px.line(response, x='month', y='total_visits', color='source',
              title='Monthly Traffic Volume by Source',
              markers=True, line_shape='spline')

# Customize the layout
fig.update_layout(xaxis_title='Month', yaxis_title='Total Visits', legend_title='Source')

# Show the interactive plot
fig.show()
```


Monthly Traffic Volume by Source
Total Visits
50k
40k
30k
20k
10k
0
Jan 2017
Feb 2017
Mar 2017
Apr 2017
May 2017
Jun 2017
Jul 2017
Month
centrum.cz
m.yz.sm.cn
meetup.com
mail.aol.com
google.com.pe
moodle.aurora.edu
wap.sogou.com
m.sp.sm.cn
amp.reddit.com
calendar.google.com
earth.google.com
news.ycombinator.com
mx.search.yahoo.com
images.google.com.au
newclasses.nyu.edu
arstechnica.com
google.bg
it.pinterest.com