



Wikipedia User Behavior Analysis

An in-depth analysis of Wikipedia's user interaction data.

Presented by Clinton Ukachukwu



Introduction

- **Wikipedia serves millions of users daily, offering insights into a wide array of topics.**
- **Understanding the patterns and trends in user interaction provides a window into the collective interests and behaviors of a global audience and offers valuable data for strategic enhancements to the platform.**
- **The objective is to uncover underlying trends and user preferences that can inform content strategy, website design, and overall user experience.**



Data

- **The dataset contains the daily number of webpage visits for 1500 Wikipedia webpages with around 549,000 entries.**
- **The dataset was sourced from Kaggle in its wide format.**
- **The data was transformed from wide to long with 3 columns for date, page names and page visits.**

Solution Approach

- “Unpivot” the table from wide to long format.

	A	B	C	D	E	
1	Page	2016-01-01	2016-01-02	2016-01-03	2016-01-04	####
2	.xxx_en.wikipedia.org_all-access_all-agents	7089	7592	7436	7032	
3	.xxx_en.wikipedia.org_mobile-web_all-agents	6182	6447	6437	5912	
4	1._Juli_de.wikipedia.org_desktop_all-agents	20	49	33	49	
5	2014_FIFA_World_Cup_en.wikipedia.org_all-access_all-agents	5363	6280	6297	6288	

A	B	C	D
	date	page	visits
0	2016-01-01 00:00:00	.xxx_en.wikipedia.org_all-access_all-agents	7089
1	2016-01-01 00:00:00	.xxx_en.wikipedia.org_mobile-web_all-agents	6182
2	2016-01-01 00:00:00	1._Juli_de.wikipedia.org_desktop_all-agents	20
3	2016-01-01 00:00:00	2014_FIFA_World_Cup_en.wikipedia.org_all-access_all-agents	5363
4	2016-01-01 00:00:00	2014_FIFA_World_Cup_en.wikipedia.org_desktop_all-agents	3186
5	2016-01-01 00:00:00	2014_FIFA_World_Cup_en.wikipedia.org_mobile-web_all-agents	2136
6	2016-01-01 00:00:00	2015_Copa_Am�rica_en.wikipedia.org_all-access_all-agents	1160
7	2016-01-01 00:00:00	2015_Copa_Am�rica_en.wikipedia.org_desktop_all-agents	564



Solution Approach

- Inspecting, understanding and cleaning the data in Python Pandas.
- Transforming the cleaned data into a SQLite database for querying.
- Exporting query results into CSV file for visualization in Tableau.
- Extracting language codes and device types in Pandas for visualization in Tableau.
- Data Analysis to answer project questions.



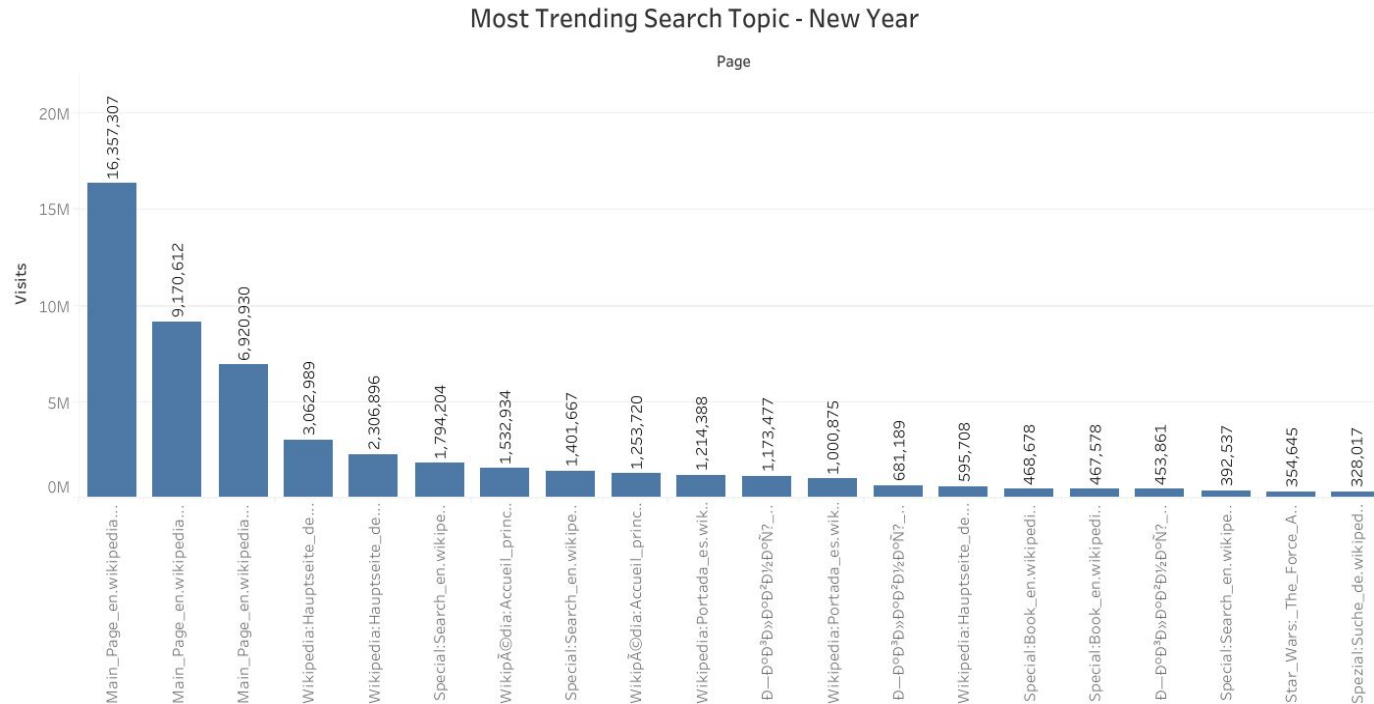
Project Questions

- **What were some of the most trending search topics on Wikipedia on the following days?**
 - **New year's day**
 - **November 8, 2016**
- **Which day(s) of the week is/are the most/least popular for visiting wikipedia?**
- **How many languages are represented in this dataset? What proportion of the pages does each language represent?**
- **Which device type is used more frequently for visiting wikipedia i.e. desktop or mobile devices?**



Findings & Insights

Most Trending Search Topics on New Year's Day

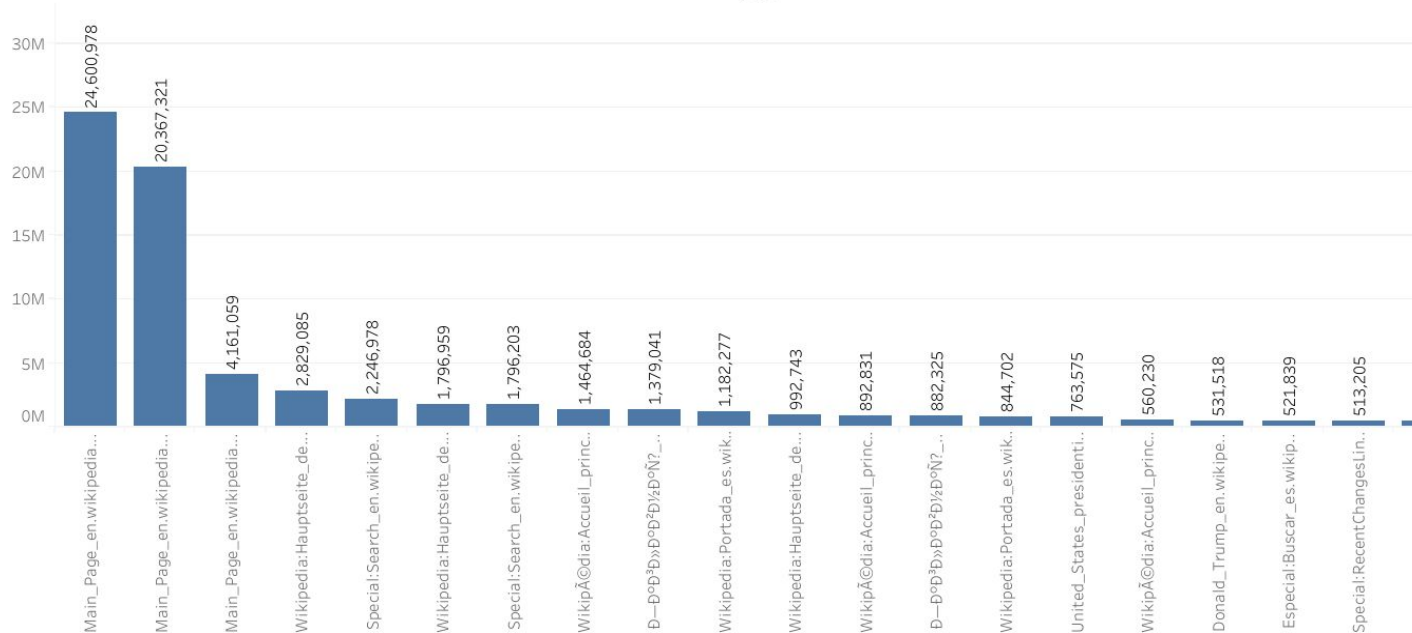


- The highest page visits was the wikipedia main page.
- The newly released Star Wars movie “The Force Awakens” made it to the top 20 most visited pages.

Most Trending Search Topics on New Year's Day

Most Trending Search Topic on Election Day

Page

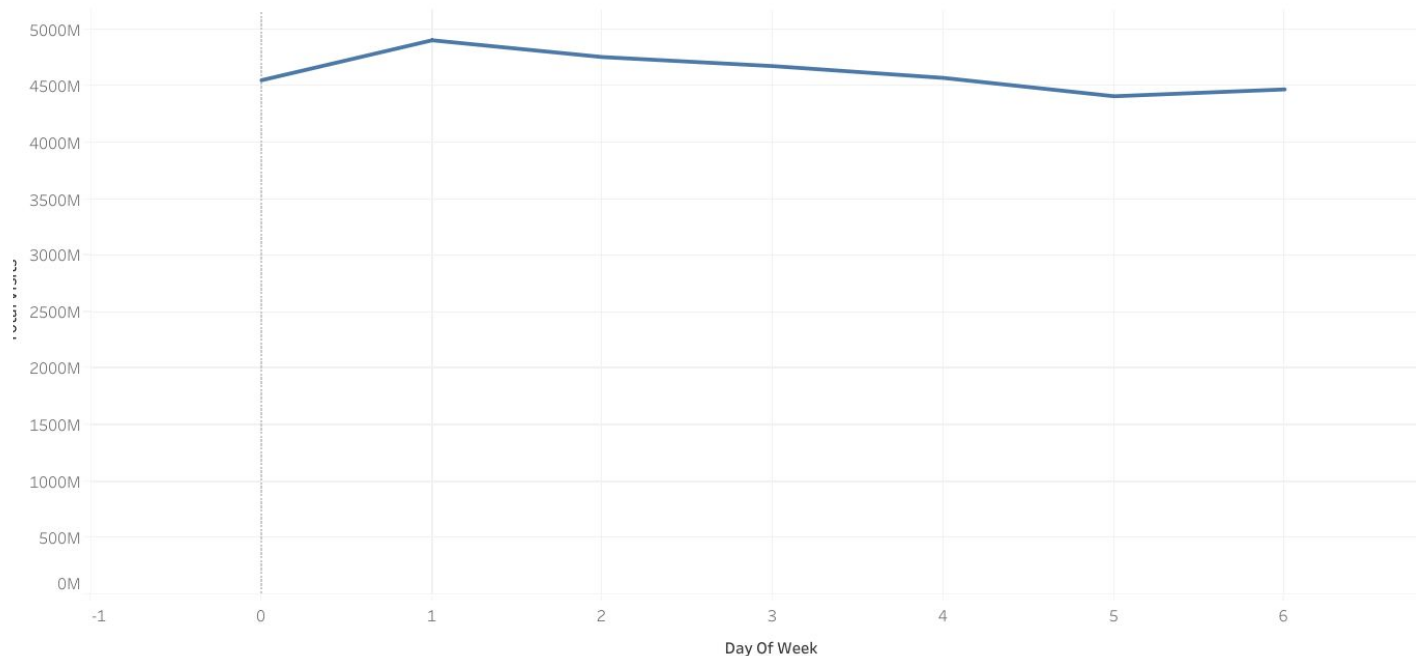


- The highest page visits was the wikipedia main page.
- The United States Presidential Election and Donald Trump appeared in the top 20 most visited pages on that day.

Popularity by Day of the Week

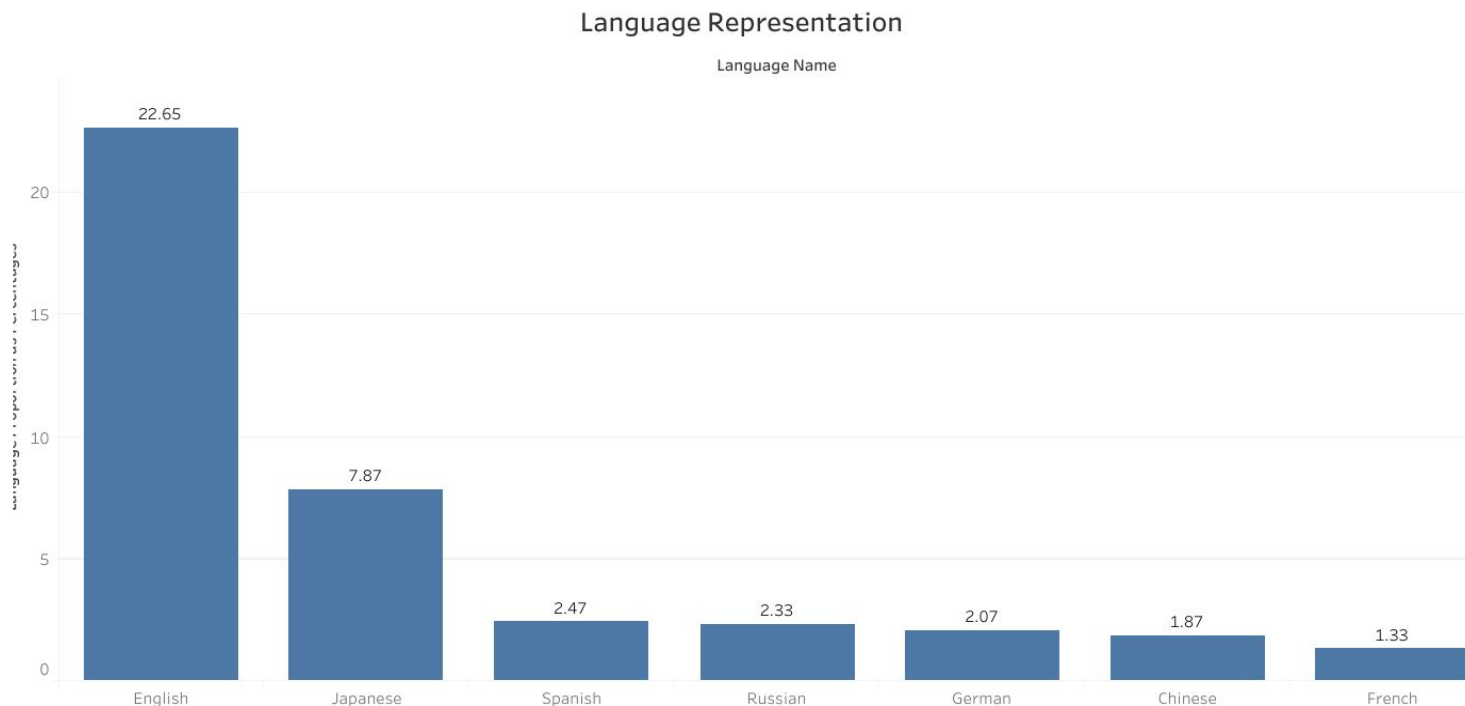


Most Popular Day of the Week



- Mondays are the most popular for visiting Wikipedia with 4.9m visits.
- Fridays are the least popular with 4.4m visits.

Language Representation

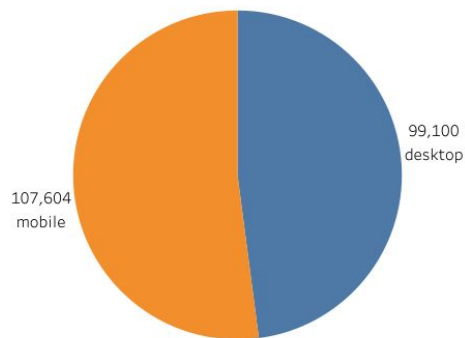


- Languages include Chinese, English, French, German, Japanese, Russian and Spanish
- English had the highest at 22.65%, followed by Japanese at 7.87% and French at 1.33%.

Device Type Usage



Most Frequently Used Device Type



- In 2016, Wikipedia had more visits from mobile devices at 107.6k visits than desktop devices at 99.1k visits.



Summary



Summary of findings

- **Most Trending Search Topics**

- The Star Wars movie, “The force awakens” trended on New Year’s day.
- The United States Presidential Election and Donald Trump trended on the day of the US elections.

- **Popularity by Day of the Week**

- Mondays are the peak of Wikipedia visits and Fridays are the least



Summary of findings

- **Language Representation**

- English and Japanese are leading as the most represented languages.
- Chinese and French are lagging as the least represented languages.

- **Device Type Usage**

- Mobile devices are the most frequently used compared to desktop devices.



Actions & Recommendations



Recommended Actions

- **Content release and updates**

- Prioritize content that provides timely information on current societal interests in the content strategy.
- Strategically schedule the release of new content, updates, or maintenance activities taking days with highest traffic into account.
- As mobile usage continues to grow, focusing on mobile-friendly content and design will be crucial for maintaining and expanding Wikipedia's user base.

- **Accessibility and inclusion**

- Enhance content in underrepresented languages to improve accessibility and inclusivity, fostering a more diverse user base.



Thank you!