Practicum –Data collection, Cleaning and Analysis using Google Trends data

Certification Programme in Business Analytics, Batch 8

Apr 2, 2017

- Data Collection: Why, what, source, how?

- Google Trends data to understand popularity through search frequency
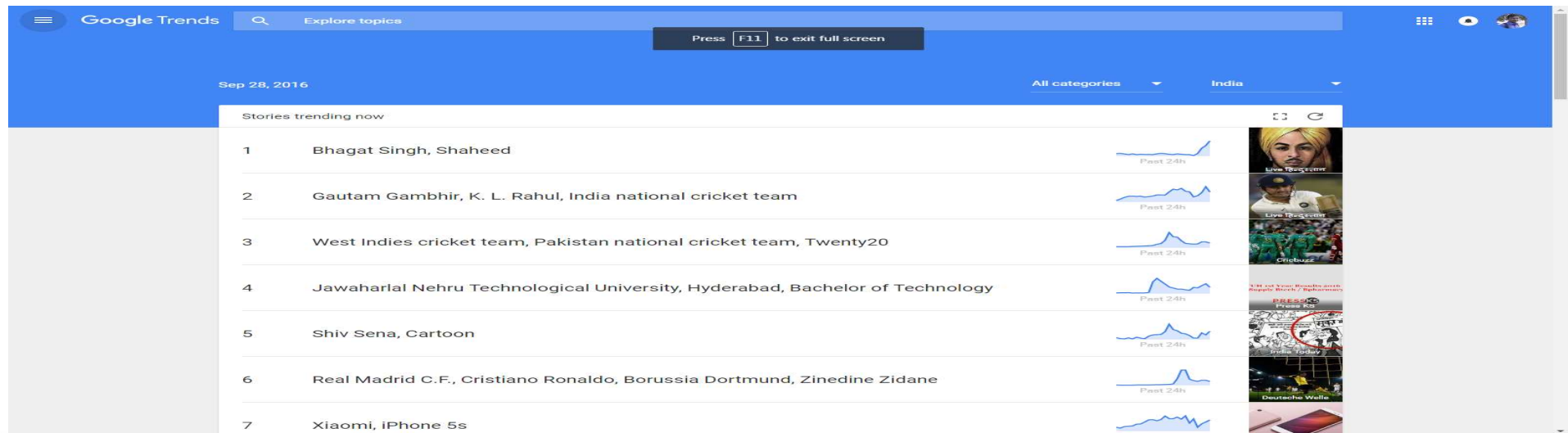
     www.google.com/trends

- Help link

     https://support.google.com/trends#topic=6248052

Google Trends

- Search Frequency, Comparison
- Filters by location, time period, categorical search
- Export the data
- Citation: "Data source: Google Trends (www.google.com/trends)"

- Examples:
  - India Pakistan
  - "India Pakistan" or "Pakistan India"
  - India+Pakistan
  - India-Pakistan or Pakistan-India
  - India+Bharat+Hindustan

Interest over time

- Search interest relative to the highest point for the given region and time.
- 100 -> peak popularity
- 50 -> half popular
- 0 -> less than 1% as popular as the peak.

India
Search term
Worldwide , Past 5 years

Pakistan
Search term
Worldwide , Past 5 years

+ Add comparison

All categories ▼   Web Search ▼

Interest over time ❓

100
75
50
25

Average   Oct 2, 2011   Jun 23, 2013   Mar 15, 2015

Interest by subregion/ city

- Ranked by each region during the specified time frame.
- Scaled from 0 to 100,
- 100 -> region with peak popularity
- 50 -> region where the term is half as popular,
- 0 -> region with less than 1% popularity.

Interest by Region / City

- Location-wise popularity during the specified time frame.
- Scale from 0 to 100,
- 100 -> location with the most popularity as a fraction of total searches
- 50 -> a location which is half as popular
- 0 -> a location where the term was less than 1% as popular as the peak.

Note: A higher value means a higher proportion of all queries, not a higher absolute query count. So a tiny country where 80% of the queries are for "bananas" will get twice the score of a giant country where only 40% of the queries are for "bananas".
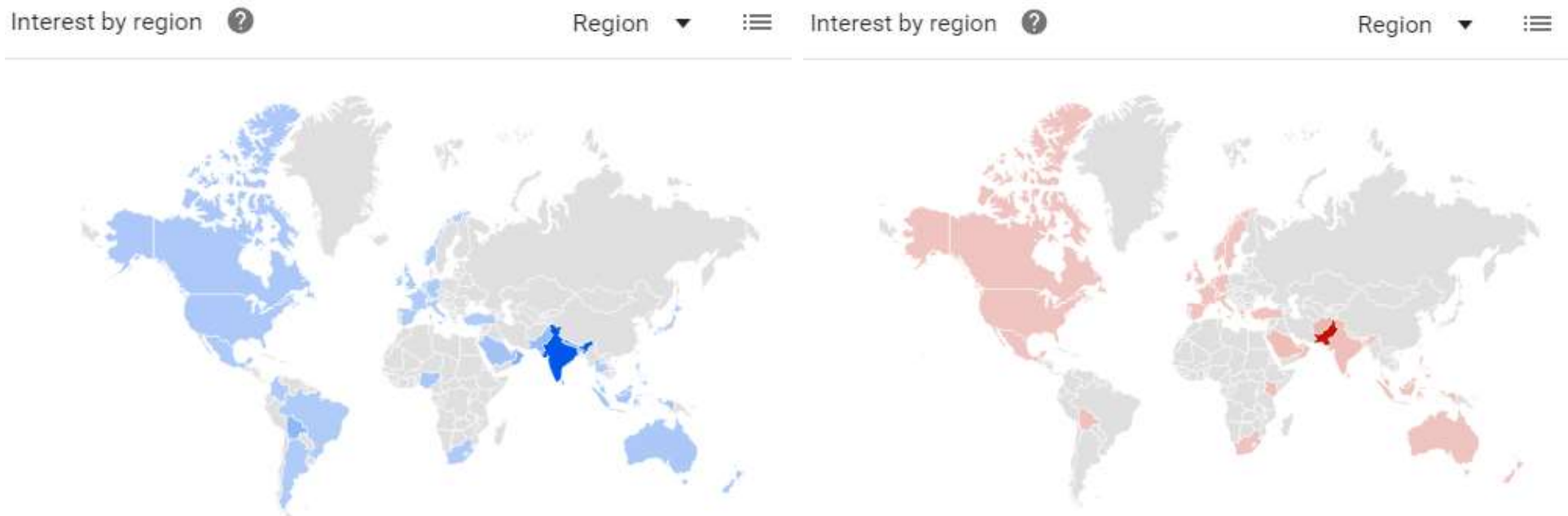
Related topics or queries (Top / Rising)

- Users searching for this term also searched for these queries.
  - Can be used in recommendation engine.
- Top - The most popular search queries.
  - 100 -> the most commonly searched query
  - 0 -> query searched for less than 1% as often as the most popular query.
- Rising - Queries with the biggest increase in search frequency since the last time period.
- "Breakout"-> 5000% (High numerator or low denominator)

Related queries	Rising
1 savdhaan india	Breakout
2 jabong	Breakout
3 rio olympics india	Breakout
4 shopclues	Breakout
5 inna india	+4,000%

Related queries	Rising
1 patna se pakistan	Breakout
2 samsung j5	Breakout
3 jeeto pakistan	Breakout
4 unblock youtube	Breakout
5 huawei p8 lite	Breakout

PROJECT DETAILS

Topics (Pick one of your choice)

1. **Analyse State (UP, Punjab, Goa, UK) Elections using Google Trend.**
    Election Commission website:
    http://www.indiavotes.com/pc/info?eid=16&state=0   http://www.indiavotes.com/pc/acwisedetails/61/7767/16

2. **Impact of Competitors' product launch**
    How do people react to launch of new product (smartphone device) by same company or competitors.
    Analyse the change in interest level of a particular product due to other, identify competitive products, timing of the launch etc.

3. **US Migration policy (H1-B, Student, Migrants, Mexico, Gulf countries etc.)**
    Interests level of people across geographies on US migration policy reform. What is the change in migration trend from different countries. Analyse the relationship between US Unemployment rate and migration. https://www.dhs.gov/immigration-statistics
    http://www.migrationpolicy.org/programs/data-hub/us-immigration-trends#source

4. **Foreign vs Indian player in IPL**
    Analyze the interest in foreign vs Indian player across the time, team, geography. Identify the important players from all the team and analyse the performance (batting/bowling/fielding) and interest levels among them. Who gets more attention- bowler, batsman, fielder, star player, captain, or team association.

5. **Impact of Investor attention on Stock Price and Volume.** Yahoo Finance, Google Finance
    • Does investor attention lead to abnormal volume and/or price?
    • When does information gathering happen– post event, closer to event?, and when does it have more impact?
    • By analyzing the relationship between search frequency and news announcement, we can also understand the information leakage.

1. Go through the report writing and ethics references
2. Pick up one topic of interest
3. Develop idea and define business problem clearly
4. Download the data
5. Use the concepts learned in first five sessions
6. Analyse the data and summarize your result
7. Focus on thought process, writing quality, and ethics.
8. Use reference material (at least 1 book, 1 web article, 1 research study, 1 multimedia source)

THANK YOU