1) Join/Combine all datasets from 2017-11-01 to 2017-11-03.

2) Find all references to Donald Trump. That is the only data important to us.

3) Generate % of accounts tweeting about Donald Trump of the total data.

4) Order accounts by the frequency of tweets about Donald Trump.

5) Find % of Donald Trump tweets which are positive in nature. (positive keywords - Good, very good, nice. We leave it to you to add more positive keywords. )

6) Find % of accounts with more than 50% positive tweets about Donald Trump.

7) Create distribution of all donald trump tweets for each hour of day, and compare standard deviation of the distributions across different hours of day. You can plot this using any tool you want.

8) Overlay hour of day distributions with accounts with more than 50% positive tweets. See if you can use this overlay to identify timezones of USA who are more supportive of Donald Trump.

Answers:

1.

**SELECT \***

**FROM `2017\_11\_01`**

**UNION ALL**

**SELECT \***

**FROM `2017\_11\_02`**

**UNION ALL**

**SELECT \***

**FROM `2017\_11\_03`**

2. **SELECT \***

**FROM**

**(**

**SELECT \***

**FROM `2017\_11\_01`**

**UNION ALL**

**SELECT \***

**FROM `2017\_11\_02`**

**UNION ALL**

**SELECT \***

**FROM `2017\_11\_03`**

**) t**

**WHERE text like '%trump%'**

3.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **SELECT** | |  |  |  |  |
| **(** |  |  |  |  |  |
| **(SELECT count(Distinct user\_id)** | | | |  |  |
| **from** |  |  |  |  |  |
| **(** |  |  |  |  |  |
| **SELECT \* from `2017\_11\_01` where text like '%trump%'** | | | | | |
| **UNION ALL** | |  |  |  |  |
| **SELECT \* from `2017\_11\_02` where text like '%trump%'** | | | | | |
| **UNION ALL** | |  |  |  |  |
| **SELECT \* from `2017\_11\_03` where text like '%trump%'** | | | | | |
| **) t )** |  |  |  |  |  |
| **/** |  |  |  |  |  |
| **(SELECT COUNT(user\_id)** | | |  |  |  |
| **FROM** |  |  |  |  |  |
| **(** |  |  |  |  |  |
| **SELECT \* from `2017\_11\_01`** | | | |  |  |
| **UNION ALL** | |  |  |  |  |
| **SELECT \* from `2017\_11\_02`** | | | |  |  |
| **UNION ALL** | |  |  |  |  |
| **SELECT \* from `2017\_11\_03`** | | | |  |  |
| **)y ) \*100** |  |  |  |  |  |
| **)** |  |  |  |  |  |

4.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **SELECT user\_id, COUNT(\*)** | | | |  |  |
| **from** |  |  |  |  |  |
| **(** |  |  |  |  |  |
| **SELECT \* from `2017\_11\_01` where text like '%trump%'** | | | | | |
| **UNION ALL** | |  |  |  |  |
| **SELECT \* from `2017\_11\_02` where text like '%trump%'** | | | | | |
| **UNION ALL** | |  |  |  |  |
| **SELECT \* from `2017\_11\_03` where text like '%trump%'** | | | | | |
| **) t** |  |  |  |  |  |
| **GROUP BY USER\_ID** | | |  |  |  |
| **ORDER BY COUNT(user\_id) DESC** | | | |  |  |

5.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **SELECT** | |  |  |  |  |  |
| **(** |  |  |  |  |  |  |
| **SELECT COUNT(\*)** | | |  |  |  |  |
| **from** |  |  |  |  |  |  |
| **(** |  |  |  |  |  |  |
| **SELECT \* from `2017\_11\_01` where text like '%trump%'** | | | | | | |
| **UNION ALL** | |  |  |  |  |  |
| **SELECT \* from `2017\_11\_02` where text like '%trump%'** | | | | | | |
| **UNION ALL** | |  |  |  |  |  |
| **SELECT \* from `2017\_11\_03` where text like '%trump%'** | | | | | | |
| **) t** |  |  |  |  |  |  |
| **where text like '%good%' or ( text like '%nice%' or text like '%happy%')** | | | | | | |
| **)/** |  |  |  |  |  |  |
| **(SELECT COUNT(\*)** | | |  |  |  |  |
| **from** |  |  |  |  |  |  |
| **(** |  |  |  |  |  |  |
| **SELECT \* from `2017\_11\_01` where text like '%trump%'** | | | | | | |
| **UNION ALL** | |  |  |  |  |  |
| **SELECT \* from `2017\_11\_02` where text like '%trump%'** | | | | | | |
| **UNION ALL** | |  |  |  |  |  |
| **SELECT \* from `2017\_11\_03` where text like '%trump%'** | | | | | | |
| **) y** |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| **)\*100** |  |  |  |  |  |  |

7. 