OPERATION ANALYTICS AND INVESTIGATING METRIC SPIKE

- **Project Description** – Find the valuable insights that can help the company's operations and understand sudden changes in key metrics.

- **Approach** –

  - Database creation – The database is created, using SQL queries on SQL Workbench.
  - Importing Data – Import the data from CSV files into the database, using SQL queries on SQL Workbench.
  - Insights extraction – Generating the insights from the database using SQL queries on SQL Workbench.

- **Tech-Stack Used** – Using SQL Workbench for this project.

- **Insights** –

  CASE STUDY 1

  - **Jobs Reviewed Over Time –**

```
SELECT COUNT(DISTINCT job_id)/(30*24) as `Job reviewed per hour`
FROM job_data WHERE ds BETWEEN '2020-11-01' AND '2020-11-30';
```

The number of jobs reviewed per hour for each day in November 2020.

	Job reviewed per hour
▶	0.0083

  - **Throughput Analysis –**

```
SELECT ds as `Date of review`,`Job reviewed`, AVG(`Job reviewed`)
OVER(ORDER BY ds ROWS BETWEEN 6 PRECEDING AND CURRENT ROW) as `Throughput 7 Rolling Average`
FROM ( SELECT ds, COUNT(DISTINCT job_id) as `Job reviewed` From job_data GROUP BY ds ORDER BY ds) a;
```

The 7-day rolling average of throughput.

Date of review	Job reviewed	Throughput 7 Rolling Average
2020-11-25	1	1.0000
2020-11-26	1	1.0000
2020-11-27	1	1.0000
2020-11-28	2	1.2500
2020-11-29	1	1.2000
2020-11-30	2	1.3333

- **Language Share Analysis -**

```
SELECT distinct language, ((COUNT(language)/(SELECT COUNT(*) FROM job_data))*100) as `Percentage share`
FROM job_data group by language;
```

The percentage share of each language over the last 30 days.

language	Percentage share
English	12.5000
Arabic	12.5000
Persian	37.5000
Hindi	12.5000
French	12.5000
Italian	12.5000

- **Duplicate Rows Detection –**

```
SELECT * FROM
( SELECT *, ROW_NUMBER() OVER( PARTITION BY job_id) as `Row Num`
FROM job_data) a WHERE `Row Num` >1;
```

Display duplicate rows from the table.

job_id	actor_id	event	language	time_spent	org	ds	Row Num
23	1005	transfer	Persian	22	D	2020-11-28	2
23	1004	skip	Persian	56	A	2020-11-26	3

CASE STUDY 2

- **Weekly User Engagement** –

```
-- TASK A
SELECT EXTRACT(WEEK FROM occurred_at) AS `Week Number`,
COUNT(DISTINCT user_id) AS `Number of Users`
FROM events GROUP BY `Week Number`;
```

The weekly user engagement.

Week Number	Number of Users
17	663
18	1068
19	1113
20	1154
21	1121
22	1186
23	1232
24	1275
25	1264
26	1302
27	1372
28	1365
29	1376
30	1467
31	1299
32	1225
33	1225
34	1204
35	104

- **User Growth Analysis –**

```
-- TASK B
SELECT `Year Num`, `Week Num`, `Num of Active Users`, SUM(`Num of Active Users`)
OVER(ORDER BY `Year Num`,`Week Num`) AS `Cummulative Active Users`
FROM (SELECT EXTRACT(Year FROM activated_at) AS `Year Num`,
EXTRACT(WEEK FROM activated_at) AS `Week Num`,
COUNT(DISTINCT user_id) AS `Num of Active Users`
FROM users WHERE state = 'active' GROUP BY `Year Num`, `Week Num`
ORDER BY `Year Num`, `Week Num`) a;
```

The user growth for the product.

Year Num	Week Num	Num of Active Users	Cummulative Active Users
2013	0	23	23
2013	1	30	53
2013	2	48	101
2013	3	36	137
2013	4	30	167
2013	5	48	215
2013	6	38	253
2013	7	42	295
2013	8	34	329
2013	9	43	372
2013	10	32	404
2013	11	31	435
2013	12	33	468
2013	13	39	507
2013	14	35	542
2013	15	43	585
2013	16	46	631
2013	17	49	680
2013	18	44	724
2013	19	57	781
2013	20	39	820
2013	21	49	869
2013	22	54	923

Year Num	Week Num	Num of Active Users	Cummulative Active Users
2013	23	50	973
2013	24	45	1018
2013	25	57	1075
2013	26	56	1131
2013	27	52	1183
2013	28	72	1255
2013	29	67	1322
2013	30	67	1389
2013	31	67	1456
2013	32	71	1527
2013	33	73	1600
2013	34	78	1678
2013	35	63	1741
2013	36	72	1813
2013	37	85	1898
2013	38	90	1988
2013	39	84	2072
2013	40	87	2159
2013	41	73	2232
2013	42	99	2331
2013	43	89	2420
2013	44	96	2516
2013	45	91	2607

Year Num	Week Num	Num of Active Users	Cummulative Active Users
2013	45	91	2607
2013	46	88	2695
2013	47	102	2797
2013	48	97	2894
2013	49	116	3010
2013	50	124	3134
2013	51	102	3236
2013	52	47	3283
2014	0	83	3366
2014	1	126	3492
2014	2	109	3601
2014	3	113	3714
2014	4	130	3844
2014	5	133	3977
2014	6	135	4112
2014	7	125	4237
2014	8	129	4366
2014	9	133	4499
2014	10	154	4653
2014	11	130	4783
2014	12	148	4931

Year Num	Week Num	Num of Active Users	Cummulative Active Users
2014	13	167	5098
2014	14	162	5260
2014	15	164	5424
2014	16	179	5603
2014	17	170	5773
2014	18	163	5936
2014	19	185	6121
2014	20	176	6297
2014	21	183	6480
2014	22	196	6676
2014	23	196	6872
2014	24	229	7101
2014	25	207	7308
2014	26	201	7509
2014	27	222	7731
2014	28	215	7946
2014	29	221	8167
2014	30	238	8405
2014	31	193	8598
2014	32	245	8843
2014	33	261	9104
2014	34	259	9363
2014	35	18	9381

- **Weekly Retention Analysis –**

```
-- TASK C
SELECT DISTINCT user_id, COUNT(user_id) as user_num,
SUM(CASE WHEN retention_week = 1 THEN 1 ELSE 0 END) AS Per_week_retention
FROM (SELECT a.user_id, a.signup_week, b.engagement_week,
b.engagement_week - a.signup_week AS retention_week
FROM (SELECT DISTINCT user_id, EXTRACT(WEEK FROM occurred_at) AS signup_week FROM events
WHERE event_type = 'signup_flow' AND event_name = 'complete_sigbup') a LEFT JOIN
(SELECT DISTINCT user_id, EXTRACT(WEEK FROM occurred_at) AS engagement_week FROM events
WHERE event_type = 'engagement' ) b ON a.user_id = b.user_id ) d
GROUP BY user_id ORDER BY user_id;
```

The weekly retention of users based on their sign-up cohort.

user_id	user_num	Per_week_retention
11768	1	0
11770	1	0
11775	2	1
11778	3	0
11779	5	1
11780	2	1
11785	1	0
11787	3	1
11791	2	1
11793	6	1
11795	2	1
11798	6	1
11799	10	1
11801	2	1
11804	2	1
11806	1	0
11809	1	0
11811	2	1
11813	6	0
11816	3	0
11818	2	1
11820	4	1

There are more than 3000 rows that's why I'm attaching the result sheet link.

- **Weekly Engagement Per Device –**

```
-- TASK D
SELECT EXTRACT(YEAR FROM occurred_At) AS year_num,
EXTRACT(WEEK FROM occurred_at) AS week_num, device,
COUNT(DISTINCT user_id) AS no_of_users from events
WHERE event_type = 'engagement' GROUP BY 1,2,3 ORDER BY 1,2,3;
```

The weekly engagement per device is-

year_num	week_num	device	no_of_users
2014	17	acer aspire desktop	9
2014	17	acer aspire notebook	20
2014	17	amazon fire phone	4
2014	17	asus chromebook	21
2014	17	dell inspiron desktop	18
2014	17	dell inspiron notebook	46
2014	17	hp pavilion desktop	14
2014	17	htc one	16
2014	17	ipad air	27
2014	17	ipad mini	19
2014	17	iphone 4s	21
2014	17	iphone 5	65
2014	17	iphone 5s	42
2014	17	kindle fire	6
2014	17	lenovo thinkpad	86
2014	17	mac mini	6
2014	17	macbook air	54
2014	17	macbook pro	143
2014	17	nexus 10	16
2014	17	nexus 5	40
2014	17	nexus 7	18

The result file link.

- **Email Engagement Analysis –**

```
-- TASK E
SELECT 100*SUM(CASE WHEN email_cat = 'email_opened' THEN 1 ELSE 0 END)/
SUM(CASE WHEN email_cat = 'email_sent' THEN 1 ELSE 0 END) AS email_opening_rate,
100*SUM(CASE WHEN email_cat = 'email_clicked' THEN 1 ELSE 0 END)/
SUM(CASE WHEN email_cat = 'email_sent' THEN 1 ELSE 0 END) AS email_clicking_rate
FROM ( SELECT *, CASE WHEN action in('sent_weekly_digest','sent_reengagement_email')
 THEN 'email_sent' WHEN action in ('email_open') THEN 'email_opened'
 WHEN action in ('email_clickthrough') THEN 'email_clicked' END as email_cat
 FROM email_events ) a;
```

The email engagement metrics.

	email_opening_rate	email_clicking_rate
▶	33.5834	14.7899

➢ **Results** –

- Fewer than 0.01 jobs were reviewed per hour throughout the month of November.
- To calculate the throughput, we will use a 7-day rolling average. The 7-day rolling average provides the mean for each day over the previous seven days, from day 1 to day 7. Whereas, the daily metric only gives the average for that specific day.
- The Persian language had the largest share among all other languages.
- User engagement was highest in the 28th week and lowest in the 17th week.
- The 34th week of 2014 had the highest number of active users. User growth increased from 23 in the first week of 2013 to 9,381 by the 35th week of 2014.
- The majority of retained users were only retained for a week, after which retention rates dropped.
- Users with the highest engagement with the product were using Macbook Pro devices.
- The email open rate is 33.58%, and the click rate is 14.78% of all the emails.