

Day31 - March7th 2024

1. Started my day waking up early and completed my activities
2. Prayed to god as it was Thursday
3. Solved 1 leetcode medium problem
4. Worked on pyspark project and completed some transformation on the data

The screenshot displays a development environment with three main components:

- YouTube Video (Left):** A video titled "Real Time Application || End-to-End Project || Part-2" is playing. The video player shows a schema for dataframes and a table of data.
- Code Editor (Middle):** The editor shows the `data_processing.py` file. The code includes logging, dataframe selection, and dropping null values. The terminal output shows the execution of the code, including a warning about dropping columns and the execution of `df_presc_sel = df_presc_sel.dropna(subset='presc_id')`.
- Terminal (Right):** The terminal shows the execution of the code, including a warning about dropping columns and the execution of `df_presc_sel = df_presc_sel.dropna(subset='presc_id')`. The output shows the execution of the code, including a warning about dropping columns and the execution of `df_presc_sel = df_presc_sel.dropna(subset='presc_id')`.

The terminal output shows the execution of the code, including a warning about dropping columns and the execution of `df_presc_sel = df_presc_sel.dropna(subset='presc_id')`.

```
6 logging.config.fileConfig('Properties/configuration/logging.config')
7
8 loggers = logging.getLogger('data_processing')
9
10 usage
11 def data_clean(df1, df2):
12     try:
13         loggers.warning('data_clean() started... ')
14
15         loggers.warning('selecting required columns and '
16                         'converting some cols into UpperCase..')
17
18         df_city_sel = df1.select(upper(col('city')).alias('city'), df1.state_id, upper(df1.state_name).alias('state_name'), upper(df1.county_name).alias('county_name'), df1.population, df1.zip)
19         loggers.warning("working on OLTP dataset and selecting couple of columns and renaming...")
20
21         df_presc_sel = df2.select(df2.npi.alias('presc_id'), df2.nppes_provider_last_org_name.alias('presc_lname'),
22                                 df2.nppes_provider_first_name.alias('presc_fname'), df2.nppes_provider_city.alias('presc_city'),
23                                 df2.nppes_provider_state.alias('presc_state'), df2.specialty_description.alias('presc_specit'),
24                                 df2.drug_name, df2.total_claim_count.alias('tx_cnt'), df2.years_of_exp, df2.total_day_supply,
25                                 df2.total_drug_cost)
26
27         loggers.warning('Adding a new column to df_presc_sel')
28         df_presc_sel = df_presc_sel.withColumn('Country_name', lit('USA'))
29
30         loggers.warning('converting Years_of_exp string to Int and replacing = ')
31         df_presc_sel = df_presc_sel.withColumn('years_of_exp', regexp_replace(col('years_of_exp'), pattern=r"= ", replacement=" "))
32
33         df_presc_sel = df_presc_sel.withColumn('years_of_exp', col('years_of_exp').cast('int'))
34
35         loggers.warning('concat first and lname')
36         df_presc_sel = df_presc_sel.withColumn('presc_fullname', concat_ws(' ', col('presc_lname'), 'presc_fname'))
37
38         loggers.warning('now dropping presc_lname and presc_fname')
39
40     except Exception as e:
41         loggers.error(f'Exception occurred: {e}')
42
43 if __name__ == '__main__':
44     data_clean()
```

- 5.
6. Ended my day by solving one complex SQL question from Ankit's YT

SQLQuery1.sql - KAUSHI\SQLEXPRESS.master (KAUSHI\jamka (75)) - Microsoft SQL Server Management Studio

File Edit View Query Project Tools Window Help

Connect -> KAUSHI\SQLEXPRESS.master (KAUSHI\jamka (75))

```

create table call_details (
    call_type varchar(10),
    call_number varchar(12),
    call_duration int
);

insert into call_details
values ('OUT','181868',13),('OUT','2159010',8),
('OUT','2159010',178),('SMS','4153810',1),('OUT','2159010',152),('OUT','9140152',18),('SMS','4162672',1),
('SMS','9168204',1),('OUT','9168204',576),('INC','2159010',5),('INC','2159010',4),('SMS','2159010',1),
('SMS','4535614',1),('OUT','181868',20),('INC','181868',54),('INC','218748',20),('INC','2159010',9),
('INC','197432',66),('SMS','2159010',1),('SMS','4535614',1);

select * from call_details;

/* Q: Write a Sql query to determine phone numbers that satisfy below conditions
1- the numbers have both incoming and outgoing calls
2- sum of duration of outgoing calls is greater than incoming calls
*/

```

107 %

Results Messages

	call_type	call_number	call_duration
1	OUT	181868	13
2	OUT	2159010	8
3	OUT	2159010	178
4	SMS	4153810	1
5	OUT	2159010	152
6	OUT	9140152	18
7	SMS	4162672	1
8	SMS	9168204	1
9	OUT	9168204	576
10	INC	2159010	5
11	INC	2159010	4
12	SMS	2159010	1
13	SMS	4535614	1
14	OUT	181868	20
15	INC	181868	54
16	INC	218748	20
17	INC	2159010	9
18	INC	197432	66
19	SMS	2159010	1
20	SMS	4535614	1

Query executed successfully.

KAUSHI\SQLEXPRESS (16.0 RTM) KAUSHI\jamka (75) master | 00:00:00 20 rows

Ready 10°C Cloudy

Ln 14 Col 15 Ch 15 INS

SQLQuery1.sql - KAUSHI\SQLEXPRESS.master (KAUSHI\jamka (75)) - Microsoft SQL Server Management Studio

File Edit View Query Project Tools Window Help

Connect -> KAUSHI\SQLEXPRESS.master (KAUSHI\jamka (75))

```

1- the numbers have both incoming and outgoing calls
2- sum of duration of outgoing calls is greater than incoming calls
*/

with cte as(
    select call_number,
    sum(case when call_type='OUT' then call_duration end) as out_duration,
    sum(case when call_type='INC' then call_duration end) as in_duration
    from call_details
    group by call_number)
select call_number
from cte
where out_duration is not null and in_duration is not null and out_duration>in_duration

/* Explanation:
Step1 : Here first we will select call_number and its call types using case when
Step2 : Now we will use group by on call_numbers and calculate total duration of
outgoing and incoming calls of a call_number
Step3 : In the last we will select the call_number using a where filter */

```

107 %

Results Messages

	call_number
1	2159010

Query executed successfully.

KAUSHI\SQLEXPRESS (16.0 RTM) KAUSHI\jamka (75) master | 00:00:00 1 rows

Ready 10°C Cloudy

Ln 26 Col 20 Ch 20 INS