

Big Data Engineering Group Assignment

Salmoli Chandra 24PGAI0057

Meera Karamta 24PGAI0003

Kishnu Srivastava 24PGAI0042

Praveen Kumar Singh 24PGAI0068

Stream Processing

```
from pyspark.sql.functions import *  
  
delta_ratings_df = spark.readStream.format("delta")\  
    .option("maxFilesPerTrigger",10)\  
    .option("ignoreChanges","true")\  
    .load("/delta/ratings")  
  
delta_ratings_df.createOrReplaceTempView("ratings")  
  
%sql  
select * from ratings
```

Business Insights

Channels Used

```
%sql  
select channel, count(*)  
from ratings  
group by channel
```

Databricks visualization. Run in Databricks to view.

#Ratings

```
%sql  
select stars, count(*)  
from ratings  
group by stars
```

Databricks visualization. Run in Databricks to view.

Channel wise avg rating

```
%sql
select channel, avg(stars) as avg_rating
from ratings
group by channel
```

Databricks visualization. Run in Databricks to view.

Change in channel ratings over time (trend)

```
%sql
select channel, date(rating_time) as rating_date, avg(stars) as
avg_rating
from ratings
group by channel, date(rating_time)
```

Databricks visualization. Run in Databricks to view.