Caption FOR Data Engg. & Data Analysis project :

Hello linkedin community ,  
I completed my project on Data Engg & Data Analytics, where I worked on project which is a combination of Data Engg. & Analytics.   
In this project I used AWS Free Tier Account, Snowflake 1 month Free Account & for Power BI my Organization Account.  
My Project DATA FLOW diagram is Divided where,  
 1.Loading data from my local machine to AWS S3 Bucket with Creating Role in IAM with full access to S3 Bucket.  
2. After Loading Data into AWS Transfer that data into Snowflake through Data Pipeline come under Data Warehousing.  
This is about Data Engineering.  
Now Data Analyst Job is start:  
3. Then that Data stored in Snowflake is extracted from snowflake for Power BI where Data Preprocessing, Data Modeling, Data Analytics & at last Data Visualization. All this task Perform on Power BI & making reports / Dashboard.   
- At last this report is send to Manager or to higher Authorities for BI understanding. Report is build into 2 formats: a. Desktop Format, b. Mobile Format for their convenience to understand things.  
  
PROJECT PROCESS:

A. Setup AWS Account (S3 & IAM) for Data Engineering :

1. IAM(Identity & Access Management) role is created name ‘aws\_role’ with full access to S3 Bucket(AmazonS3FullAccess). & ARN is created for loading data on AWS.  
2. S3 Bucket created name ‘aws\_bucket’. Under S3 Bucket Folder is created ‘aws\_S3\_folder’. Under that folder Data Loading is performed from Local Machine.

- First Phase is Completed.

B. Pipeline through Data Flow from AWS to Snowflake Data Warehouse:

1. In Snowflake, default ‘Account Admin’ Role is created & Warehouse created ‘Compute\_WH’ with capacity ‘X-Small’.   
2. SQL sheet is created for Project.Database created name ‘P7’ & under database, Schema is created name ‘P7\_schema’. DB ‘P7’ & Schema ‘P7\_Schema’ is use for pipeline.  
3. Integration object is created name ‘a\_d’ & then describing that integrated object. Under that object S3 storage provider, external\_stage type, enabled equal true & getting ARN from ‘aws\_role’ with storage\_allowed\_locations=('s3://aws\_bucket/aws\_S3\_folder/').   
- After describing “a\_d” we get ‘STORAGE\_AWS\_IAM\_USER\_ARN’ & ‘STORAGE\_AWS\_EXTERNAL\_ID’ created. Use this ARN & Ext.ID for Trust Relationship is established between AWS & Snowflake.   
4. Creating ‘Source Table’ for getting data from AWS into Snowflake through Pipeline. Then, stream is created on source table, stream name Source\_Stream.  
5. File format created name: ‘a\_d\_f1’ for getting data in CSV form from AWS to Snowflake in format: CREATE or REPLACE FILE FORMAT a\_d\_f1  
TYPE = 'CSV'  
COMPRESSION = 'AUTO'  
FIELD\_DELIMITER =','  
RECORD\_DELIMITER = '\n'  
SKIP\_HEADER =1   
FIELD\_OPTIONALLY\_ENCLOSED\_BY = '\042'  
TRIM\_SPACE = FALSE  
ERROR\_ON\_COLUMN\_COUNT\_MISMATCH = TRUE  
ESCAPE = 'NONE'  
ESCAPE\_UNENCLOSED\_FIELD = '\134'  
DATE\_FORMAT = 'AUTO'  
TIMESTAMP\_FORMAT = 'AUTO;  
6. Stage is created name ‘s1\_stage’ for data transferring or getting data from AWS to snowflake through stage. Containing same URL=('s3://aws\_bucket/aws\_S3\_folder/') use for integration object ‘a\_d’ & file\_format = a\_d\_f1;  
- after listing stage connectionis established & data get on stage: ‘@s1\_stage’.

7. Pipeline is created name:’pipe\_a’ with auto\_ingest = true as also copy into Source Table from the stage @s1\_stage & on error= continue. After creating Pipe we have to describe the pipe ‘pipe\_a’. After describing pipe we get event notification for S3 Bucket in AWS.  
- In AWS -S3 Bucket we created a folder. Under that folder, In Properties we have to create a Event, event Name:event1. In that event we created object with all access to ‘all object create events: put, post, copy, Multipart upload completed’ & in Destination we used SQS Queue to enter ARN Notification from Pipeline ‘pipe\_a’. Pasted into that SQS queue & at last Event is created.  
- after event creation, set it to ON & also refreshed the Pipe.  
8. Target table is created which is fully similar to source table & use to get data from it.  
9. Task is created, name: ‘a\_task’. This task is used to transferring data from AWS Cloud to Snowflake WareHouse through PipeLine like a trigger.  
- Under task compute\_wh warehouse is used, schedule about 1 minute to transfer data where condition: when system$stream\_has\_data('source\_stream'). Stream in connected to source\_table, From source\_table data directly go into source\_stream & then from that source stream data transfer into target table.  
- selecting all columns from source\_stream where condition is: [where metadata$action='INSERT' and metadata$isupdate='FALSE'; ]  
- Created task is bydefault under Suspended Mode, So we have to resume Task   
command:[ alter task a\_task resume;].  
   
\* Second Phase is Completed, where data transferred from AWS to Snowflake through PipeLine & Data Engineering is completed.

C. Data Analysis through Power BI:  
1. Data Present in Snowflake in Target\_Table get extracted from it for Power BI where, where Data Preprocessing, Data Modeling, Data Analytics & at last Data Visualization. All this task Perform on Power BI & making reports / Dashboard.   
- At last this report is send to Manager or to higher Authorities for BI understanding. Report is build into 2 formats: a. Desktop Format, b. Mobile Format for their convenience to understand things.  
  
Let's analyze the HR analytics dashboards from the provided images.

### Common Metrics Across Dashboards

1. \*\*Count of Employees\*\*: The total number of employees in each department.

2. \*\*Attrition\*\*: The number of employees who have left the company.

3. \*\*Attrition Rate\*\*: The percentage of employees who have left the company.

4. \*\*Average Age\*\*: The average age of the employees.

5. \*\*Average Salary\*\*: The average salary of the employees.

6. \*\*Average Years\*\*: The average number of years employees have been with the company.

### Detailed Analysis of Each Dashboard

#### 1. \*\*First Dashboard\*\*

   - \*\*Department\*\*: Human Resources

   - \*\*Count of Employees\*\*: 446

   - \*\*Attrition\*\*: 92

   - \*\*Attrition Rate\*\*: 20.6%

   - \*\*Average Age\*\*: 37

   - \*\*Average Salary\*\*: 7.0K

   - \*\*Average Years\*\*: 7.3

   - \*\*Attrition by Education\*\*:

     - Marketing: 38%

     - Life Sciences: 32%

     - Medical: 15%

     - Technical Degree: 11%

   - \*\*Attrition by Age\*\*:

     - 26-35: 41

     - 18-25: 18

     - 46-55: 17

     - 36-45: 16

     - 55+: 0

   - \*\*Attrition by Salary Slab\*\*:

     - Up to 5k: 43

     - 5k-10k: 36

     - 10k-15k: 11

     - 15k+: 2

   - \*\*Attrition by Years at Company\*\*:

     - 0-1: 17

     - 1-2: 7

     - 2-3: 7

     - 3-4: 6

     - 4-5: 5

     - 5-10: 4

     - 10+: 2

   - \*\*Attrition by Gender\*\*:

     - Male: 47

     - Female: 35

   - \*\*Attrition by Job Role\*\*:

     - Sales Executive: 57

     - Sales Representative: 33

     - Manager: 2

#### 2. \*\*Second Dashboard\*\*

   - \*\*Department\*\*: Research & Development

   - \*\*Count of Employees\*\*: 1470

   - \*\*Attrition\*\*: 237

   - \*\*Attrition Rate\*\*: 16.1%

   - \*\*Average Age\*\*: 37

   - \*\*Average Salary\*\*: 6.5K

   - \*\*Average Years\*\*: 7.0

   - \*\*Attrition by Education\*\*:

     - Life Sciences: 38%

     - Medical: 27%

     - Technical Degree: 14%

     - Marketing: 15%

   - \*\*Attrition by Age\*\*:

     - 26-35: 116

     - 18-25: 44

     - 36-45: 43

     - 46-55: 26

     - 55+: 8

   - \*\*Attrition by Salary Slab\*\*:

     - Up to 5k: 163

     - 5k-10k: 49

     - 10k-15k: 20

     - 15k+: 5

   - \*\*Attrition by Years at Company\*\*:

     - 0-1: 16

     - 1-2: 16

     - 2-3: 9

     - 3-4: 8

     - 4-5: 12

     - 5-10: 18

     - 10+: 8

   - \*\*Attrition by Gender\*\*:

     - Male: 140

     - Female: 79

   - \*\*Attrition by Job Role\*\*:

     - Laboratory Technician: 62

     - Sales Executive: 57

     - Research Scientist: 47

     - Sales Representative: 33

     - Manager: 5

#### 3. \*\*Third Dashboard\*\*

   - \*\*Department\*\*: Human Resources

   - \*\*Count of Employees\*\*: 63

   - \*\*Attrition\*\*: 12

   - \*\*Attrition Rate\*\*: 19.0%

   - \*\*Average Age\*\*: 38

   - \*\*Average Salary\*\*: 6.7K

   - \*\*Average Years\*\*: 7.2

   - \*\*Attrition by Education\*\*:

     - Human Resources: 58%

     - Technical Degree: 17%

     - Medical: 17%

   - \*\*Attrition by Age\*\*:

     - 26-35: 8

     - 18-25: 2

     - 36-45: 2

     - 46-55: 0

     - 55+: 0

   - \*\*Attrition by Salary Slab\*\*:

     - Up to 5k: 10

     - 5k-10k: 1

     - 10k-15k: 1

     - 15k+: 0

   - \*\*Attrition by Years at Company\*\*:

     - 0-1: 4

     - 1-2: 2

     - 2-3: 2

     - 3-4: 0

     - 4-5: 1

     - 5-10: 1

     - 10+: 1

   - \*\*Attrition by Gender\*\*:

     - Male: 5

     - Female: 6

   - \*\*Attrition by Job Role\*\*:

     - Human Resources: 12

     - Manager: 0

#### 4. \*\*Fourth Dashboard\*\*

   - \*\*Department\*\*: Research & Development

   - \*\*Count of Employees\*\*: 961

   - \*\*Attrition\*\*: 133

   - \*\*Attrition Rate\*\*: 13.8%

   - \*\*Average Age\*\*: 37

   - \*\*Average Salary\*\*: 6.3K

   - \*\*Average Years\*\*: 6.9

   - \*\*Attrition by Education\*\*:

     - Life Sciences: 44%

     - Medical: 35%

     - Technical Degree: 15%

   - \*\*Attrition by Age\*\*:

     - 26-35: 67

     - 36-45: 25

     - 18-25: 24

     - 46-55: 9

     - 55+: 8

   - \*\*Attrition by Salary Slab\*\*:

     - Up to 5k: 110

     - 5k-10k: 12

     - 10k-15k: 8

     - 15k+: 3

   - \*\*Attrition by Years at Company\*\*:

     - 0-1: 9

     - 1-2: 8

     - 2-3: 8

     - 3-4: 5

     - 4-5: 6

     - 5-10: 14

     - 10+: 3

   - \*\*Attrition by Gender\*\*:

     - Male: 88

     - Female: 38

   - \*\*Attrition by Job Role\*\*:

     - Laboratory Technician: 62

     - Research Scientist: 47

     - Manufacturing Director: 10

     - Healthcare Representative: 9

### Key Takeaways

- The attrition rate varies across departments, with the highest in the Human Resources department in the first dashboard (20.6%).

- Employees aged 26-35 tend to have the highest attrition rate across all departments.

- Marketing and Life Sciences are the most common education backgrounds among those who left.

- Most attrition occurs among employees earning up to 5k.

- Job roles like Sales Executive and Laboratory Technician have higher attrition rates in their respective departments.

These insights can help the organization understand where they need to focus their retention strategies and address underlying issues causing high attrition in specific departments and job roles.