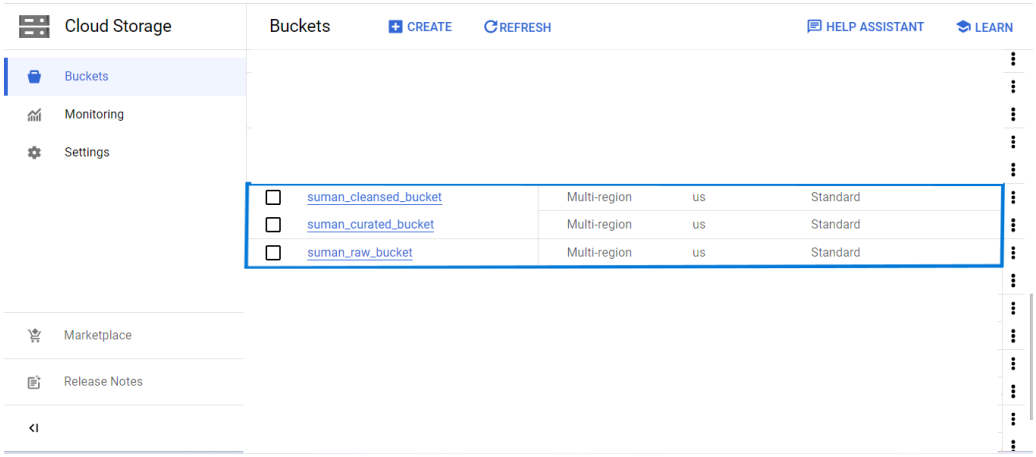
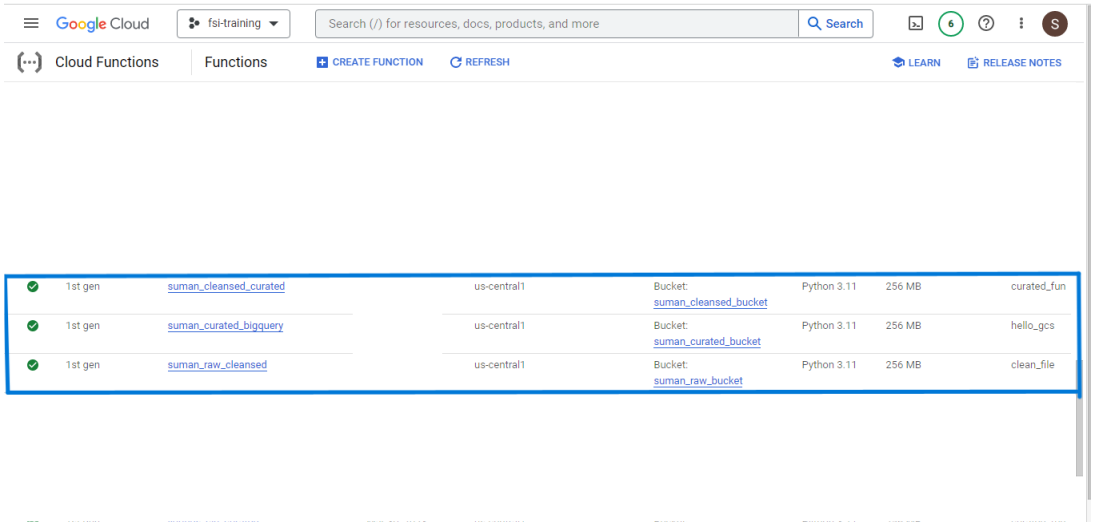
INSURANCE CLAIM ANALYSIS

Cloud Bucket



Cloud Function



suman_raw_cleansed - cloud function

```
import pandas as pd
from google.cloud import storage
import io
import pytz

def clean_file(event, context):

    file_name = event['name']

    # Create a client object
    client = storage.Client()

    # Get a reference to the bucket containing the file
    bucket = client.get_bucket('suman_raw_bucket')
```

```
    # Get a reference to the object (blob) in the bucket
    blob = bucket.blob(file_name)

    # Download the contents of the file
    content = blob.download_as_string()

    #reads a CSV file from a bytes-like object content into a pandas DataFrame df.
    df = pd.read_csv(io.BytesIO(content))
    df['date']= pd.Timestamp.today().strftime('%Y-%m-%d')
    df['date']= pd.to_datetime(df['date'])
    df['date']=df['date'].dt.tz_localize('EST')
    df['date'] = df['date'].dt.date

    mean_value= df['age'].mean()
    df['age'].fillna(value=mean_value, inplace=True)
    df['age']=df['age'].astype('int')

    df['region']=df['region'].fillna(df.region.mode()[0])

    new_file_name = "Cleansed_"+file_name
    destination_bucket=client.get_bucket("suman_cleansed_bucket")
    new_blob = destination_bucket.blob(new_file_name)
    new_blob.upload_from_string(df.to_csv(index=False), 'text/csv')
```

suman_cleansed_curated - cloud function

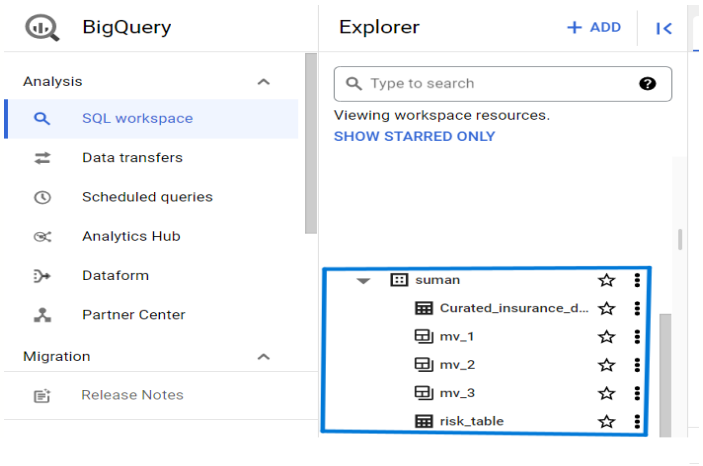
```
import pandas as pd
from google.cloud import storage
import io
import numpy as np

def curated_fun(event, context):

    file_name = event['name']
    client = storage.Client()
    bucket = client.get_bucket('suman_cleansed_bucket')
    blob = bucket.blob(file_name)
    content = blob.download_as_string()

    df = pd.read_csv(io.BytesIO(content))

    condition=[(df['bmi'] < 18.5),
    (df['bmi'] >= 18.5) & (df['bmi'] <  25),
    (df['bmi'] >= 25) & (df['bmi'] < 30),
    (df['bmi'] >= 30)]
    values = ["underweight", "healthy_weight", "overweight", "obesity"]

    df['bmi_category'] = np.select(condition, values)

    condition1=[(df['bloodpressure'] < 120),
    (df['bloodpressure'] >= 120) & (df['bloodpressure'] < 130),
    (df['bloodpressure'] >= 130) & (df['bloodpressure'] < 140),
    (df['bloodpressure'] >= 140) & (df['bloodpressure'] < 180),
    (df['bloodpressure'] >= 180)]

    values1 = ["normal", "elevated", "hbp-stage1",
```

```
    "hbp-stage2","hypertensive_crisis"]

    df['bloodpressure_category'] = np.select(condition1, values1)
    new_file_name = "Curated_insurance_data.json"
    destination_bucket=client.get_bucket("suman_curated_bucket")
    new_blob = destination_bucket.blob(new_file_name)
    new_blob.upload_from_string(df.to_json(orient = 'records'), 'application/json')
```

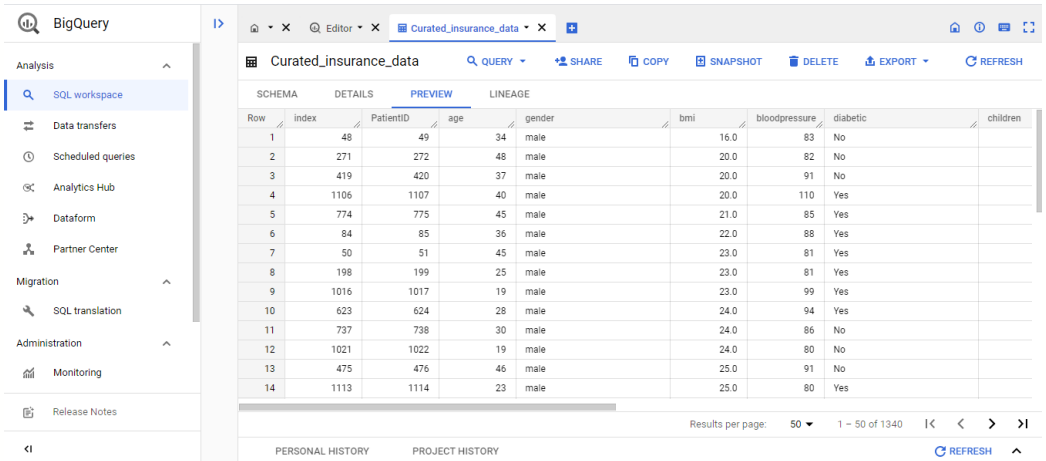
suman_curated_bigquery - cloud function

```
import pandas as pd
from pandas.io import gbq
from google.cloud import bigquery

def hello_gcs(event, context):
  file_name = event['name']
  table_name = file_name.split('.')[0]
  df_data = pd.read_json('gs://' + event['bucket'] + '/' + file_name)
  df_data.to_gbq('q-gcp-buc1-fsi-training-23-03.suman.' + table_name,
                 project_id='q-gcp-buc1-fsi-training-23-03',
                 if_exists='append',
                 location='us')
```

BigQuery




```
SELECT * FROM `q-gcp-buc1-fsi-training-23-03.suman.Curated_insurance_data` ;
```



Materialized view-1

```
create or replace materialized view q-gcp-buc1-fsi-training-23-03.suman.mv_1 as
SELECT gender,
CASE
WHEN age < 30 THEN '18-29'
WHEN age < 40 THEN '30-39'
WHEN age < 50 THEN '40-49'
ELSE '50+'
END as age_group,
AVG(claim) as avg_claim
FROM `q-gcp-buc1-fsi-training-23-03.suman.Curated_insurance_data`
GROUP BY gender, age_group;
```

Query results

SAVE RESULTS

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS | EXECUTION GRAPH PREVIEW

Row	gender	age_group	avg_claim
1	male	50+	16174.88
2	male	18-29	14874.45
3	male	40-49	13664.61
4	male	30-39	12876.66
5	female	30-39	12861.52
6	female	50+	12762.22
7	female	40-49	12548.59
8	female	18-29	11747.39

Materialized view-2

```
create or replace materialized view q-gcp-buc1-fsi-training-23-03.suman.mv_2 as
SELECT bmi_category,
bloodpressure_category,
AVG(claim) as average_claim
FROM `q-gcp-buc1-fsi-training-23-03.suman.Curated_insurance_data`
GROUP BY bmi_category, bloodpressure_category;
```

Query results

SAVE RESULTS

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS | EXECUTION GRAPH PREVIEW

Row	bmi_category	bloodpressure_category	average_claim
1	obesity	elevated	39617.57
2	obesity	hbp-stage1	39436.77
3	obesity	hbp-stage2	38323.5
4	overweight	hbp-stage1	30946.83
5	overweight	elevated	28101.33
6	healthy_weight	hbp-stage1	27206.9
7	healthy_weight	elevated	26926.51
8	obesity	normal	13677.14
9	overweight	normal	10747.77
10	healthy_weight	normal	10111.88
11	underweight	normal	8852.2

Results per page: 50 1 – 11 of 11

Materialized view-3

```
create or replace table q-gcp-buc1-fsi-training-23-03.suman.risk_table as( select
case
when t1.risk_score>=50 then 'very high'
when t1.risk_score<50 and t1.risk_score>=40 then 'high'
when t1.risk_score<40 and risk_score>=30 then 'medium'
else 'low'
end as risk_group,count(*) as num_policies,
round((sum(t1.claim)),2) as total_claim,
round((avg(t1.claim)),2) as avg_claim
from(select patientid,
round(((0.4*age)+(0.2*bloodpressure)+
(0.2*case when diabetic ='Yes' then 1 else 0 end)+
(0.2 * case when smoker='Yes' then 1 else 0 end)+(0.1*bmi)+
(0.1* case when gender='female' then 1 else 0 end)+
(0.05*children)),2) as risk_score,claim
FROM `q-gcp-buc1-fsi-training-23-03.suman.Curated_insurance_data`)t1
group by risk_group
order by
case
when risk_group='very high' then 1
when risk_group='high' then 2
when risk_group='medium' then 3
else 4
end);
create or replace materialized view q-gcp-buc1-fsi-training-23-03.suman.mv_3 as
select * from q-gcp-buc1-fsi-training-23-03.suman.risk_table;
select * from q-gcp-buc1-fsi-training-23-03.suman.mv_3 order by avg_claim desc;
```

Query results

SAVE RESULTS | E

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS | EXECUTION GRAPH PREVIEW

Row	patientid	risk_score	claim
1	49	31.8	1694.8
2	272	37.75	4005.42
3	420	35.05	5855.9
4	1107	40.4	22412.65
5	775	37.4	11013.71
6	85	34.45	1964.78
7	51	36.7	1704.57
8	199	28.7	1704.57
9	1017	30.2	17361.77
10	624	32.7	8604.48
11	738	31.6	10422.92

Results per page: 50 | 1 – 50 of 1340

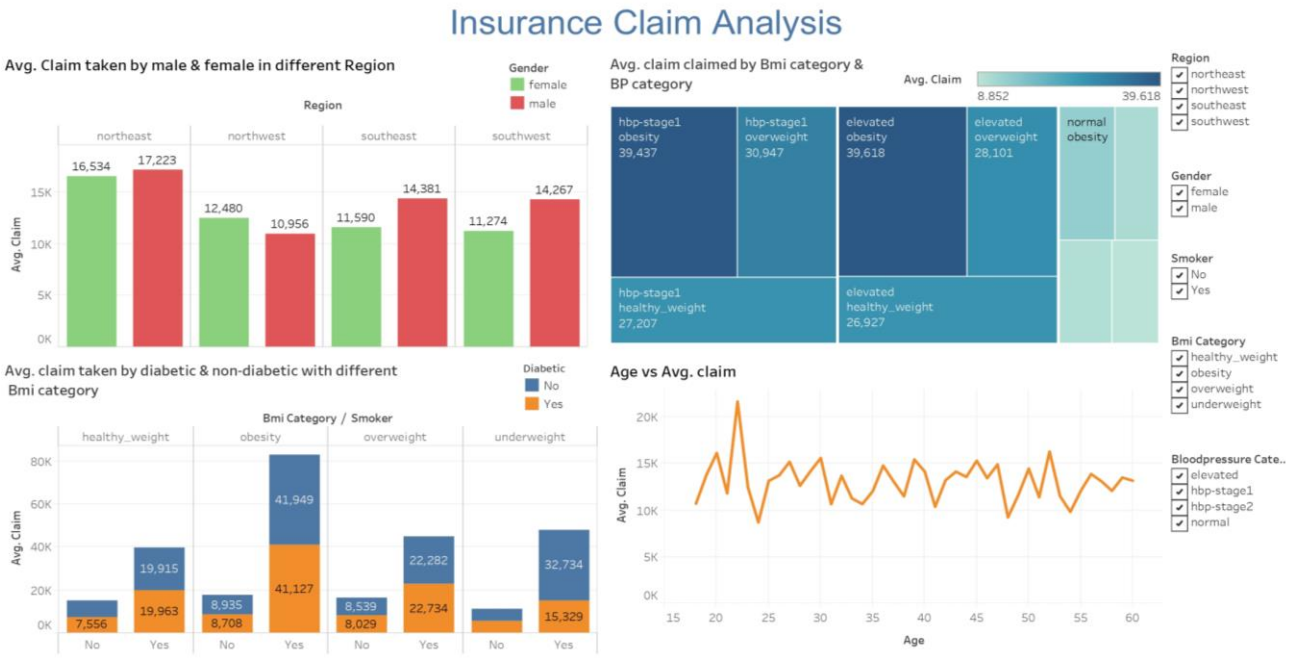
Query results

SAVE RESULTS

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS | EXECUTION GRAPH PREVIEW

Row	risk_group	num_policies	total_claim	avg_claim
1	very high	10	406401.54	40640.15
2	high	402	6562446.29	16324.49
3	medium	838	9875348.78	11784.43
4	low	90	914482.55	10160.92

Tableau


Insurance Claim Analysis
Avg. Claim taken by male & female in different Region
Gender
female
male
Region
northeast
northwest
southeast
southwest
Avg. Claim
16,534
17,223
12,480
10,956
11,590
14,381
11,274
14,267
Avg. claim claimed by Bmi category & BP category
Avg. Claim
8.852
39.618
hbp-stage1 obesity 39,437
hbp-stage1 overweight 30,947
elevated obesity 39,618
elevated overweight 28,101
normal obesity
hbp-stage1 healthy_weight 27,207
elevated healthy_weight 26,927
Region
northeast
northwest
southeast
southwest
Gender
female
male
Smoker
No
Yes
Bmi Category
healthy_weight
obesity
overweight
underweight
Bloodpressure Cate..
elevated
hbp-stage1
hbp-stage2
normal
Avg. claim taken by diabetic & non-diabetic with different Bmi category
Diabetic
No
Yes
Bmi Category / Smoker
healthy_weight
obesity
overweight
underweight
Avg. Claim
7,556
19,915
19,963
8,935
8,708
41,949
41,127
8,539
8,029
22,282
22,734
32,734
15,329
No
Yes
Age vs Avg. claim
Avg. Claim
Age