Project Title: ETL Project with Data Fusion and BigQuery And Create Dashboard

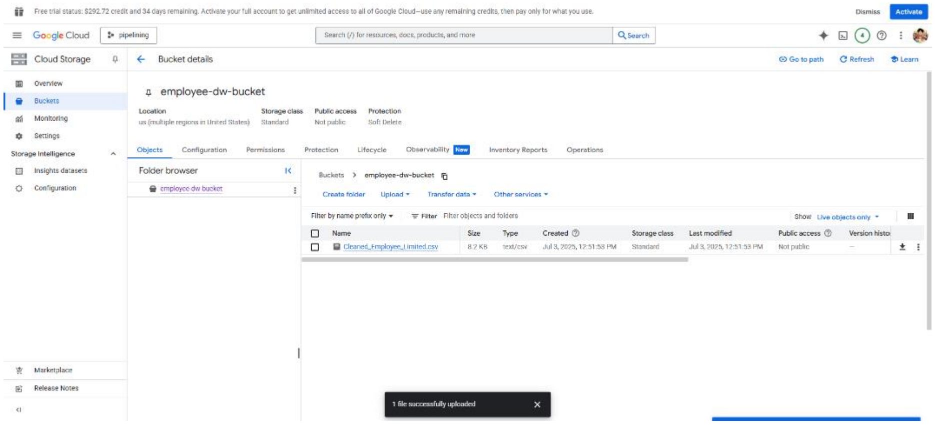
Project Overview: This project is about building a full ETL (Extract, Transform, Load) process using Google Cloud tools. The goal is to clean employee data, move it through a cloud pipeline, and create a dashboard to see useful insights. All steps are done using simple and scalable tools provided by Google Cloud Platform (GCP).

Tools Used: - Python (for data cleaning) - Google Cloud Storage (to store data) - Google Cloud Data Fusion (to create ETL pipeline) - BigQuery (to store and query final data) - Looker Studio (to create dashboard)

Steps:

1. **Data Cleaning with Python:**
    - Loaded the dataset (CSV file) using pandas
    - Removed duplicate records
    - Handled missing values
    - Limited the dataset to 200 clean rows
    - Saved the cleaned file

2. **Upload to Google Cloud Storage:**
    - Created a bucket in GCS
    - Uploaded the cleaned CSV file to the bucket



Free trial status: $292.72 credit and 34 days remaining. Activate your full account to get unlimited access to all of Google Cloud—use any remaining credits, then pay only for what you use.

employee-dw-bucket

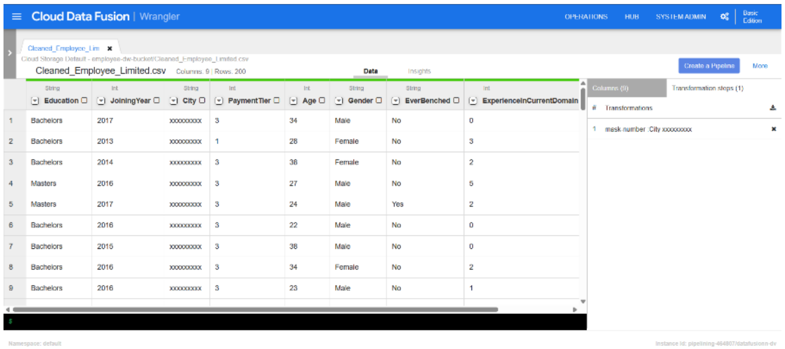
Location	Storage class	Public access	Protection
us (multiple regions in United States)	Standard	Not public	Soft Delete

Buckets > employee-dw-bucket

Name	Size	Type	Created	Storage class	Last modified	Public access	Version histo
Cleaned_Employee_Limited.csv	8.2 KB	text/csv	Jul 3, 2025, 12:51:53 PM	Standard	Jul 3, 2025, 12:51:53 PM	Not public	–

1 file successfully uploaded

3. **Create ETL Pipeline with Data Fusion:**
    - Created a new Data Fusion instance
    - Apply data masking techniques to sensitive information in Cloud Data Fusion before loading it into BigQuery.
    - Designed a pipeline to read from Cloud Storage
    - Set the pipeline to load data into BigQuery
    - Deployed the pipeline and confirmed data reached BigQuery



Cloud Storage Default - employee-dw-bucket/Cleaned_Employee_Limited.csv

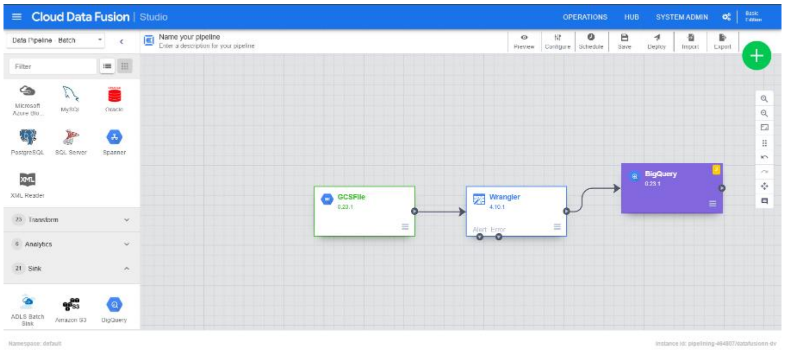
Cleaned_Employee_Limited.csv Columns: 9 | Rows: 200

	String Education	Int JoiningYear	String City	Int PaymentTier	Int Age	String Gender	String EverBenched	Int ExperienceInCurrentDomain
1	Bachelors	2017	xxxxxxxxx	3	34	Male	No	0
2	Bachelors	2013	xxxxxxxxx	1	28	Female	No	3
3	Bachelors	2014	xxxxxxxxx	3	38	Female	No	2
4	Masters	2016	xxxxxxxxx	3	27	Male	No	5
5	Masters	2017	xxxxxxxxx	3	24	Male	Yes	2
6	Bachelors	2016	xxxxxxxxx	3	22	Male	No	0
7	Bachelors	2015	xxxxxxxxx	3	38	Male	No	0
8	Bachelors	2016	xxxxxxxxx	3	34	Female	No	2
9	Bachelors	2016	xxxxxxxxx	3	23	Male	No	1

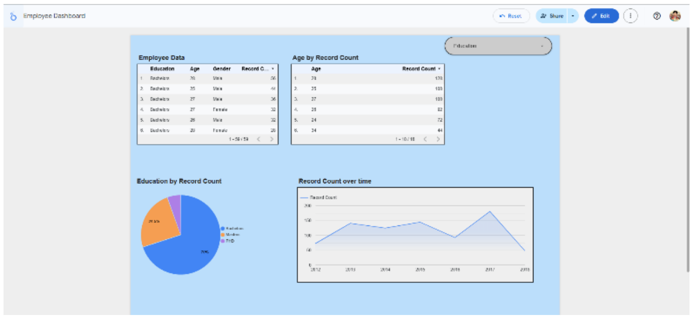
#	Transformations
1	mask-number :City xxxxxxxxx

Namespace: default

Instance id: pipelining-464807/datafusionn-dv



4. **Visualize in Looker Studio:**
    - Connected Looker Studio to BigQuery table
    - Created graphs and charts to show employee insights like age, city, etc.
    - Shared the dashboard for viewing



Employee Data

	Education	Age	Gender	Record C...
1.	Bachelors	28	Male	56
2.	Bachelors	25	Male	44
3.	Bachelors	27	Male	36
4.	Bachelors	27	Female	32
5.	Bachelors	26	Male	32
6.	Bachelors	28	Female	28

1 - 59 / 59

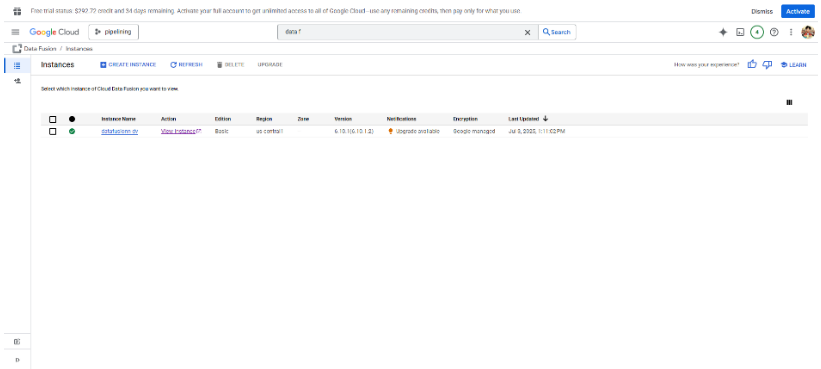
Age by Record Count

	Age	Record Count
1.	28	170
2.	25	100
3.	27	100
4.	26	82
5.	24	72
6.	34	44

1 - 10 / 18

Result: The final project is a working cloud data pipeline that starts from raw data and ends in a clean, visual dashboard. This setup can be used for HR analytics or any company working with employee records.

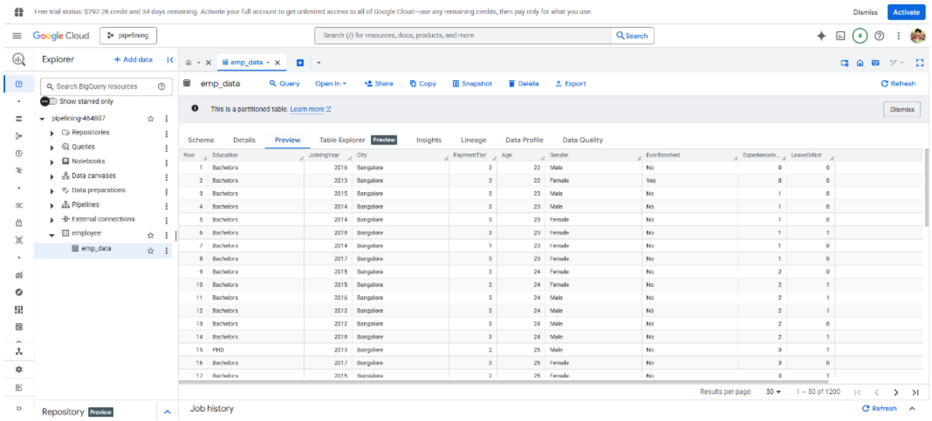
Conclusion: This project shows how to build a complete ETL and BI (Business Intelligence) system using only GCP tools and Python, which is good for real-world use.



Free trial status: $292.72 credit and 34 days remaining. Activate your full account to get unlimited access to all of Google Cloud—use any remaining credits, then pay only for what you use.

Select which instance of Cloud Data Fusion you want to view.

□	●	Instance Name	Action	Edition	Region	Zone	Version	Notifications	Encryption	Last Updated ↓
□	✓	datafusionn dv	View Instance	Basic	us central1	-	6.10.1(6.10.1.2)	Upgrade available	Google managed	Jul 8, 2025, 1:11:02PM



Free trial status: $292.26 credit and 34 days remaining. Activate your full account to get unlimited access to all of Google Cloud—use any remaining credits, then pay only for what you use.

This is a partitioned table. Learn more

Row	Education	JoiningYear	City	PaymentTier	Age	Gender	EverBenched	ExperienceIn...	LeaveOrNot
1	Bachelors	2016	Bangalore	3	22	Male	No	0	0
2	Bachelors	2013	Bangalore	3	22	Female	Yes	0	0
3	Bachelors	2015	Bangalore	3	23	Male	No	1	0
4	Bachelors	2014	Bangalore	3	23	Male	No	1	0
5	Bachelors	2014	Bangalore	3	23	Female	No	1	0
6	Bachelors	2018	Bangalore	3	23	Female	No	1	1
7	Bachelors	2014	Bangalore	1	23	Female	No	1	0
8	Bachelors	2017	Bangalore	3	23	Female	No	1	0
9	Bachelors	2015	Bangalore	3	24	Female	No	2	0
10	Bachelors	2015	Bangalore	3	24	Female	No	2	1
11	Bachelors	2016	Bangalore	3	24	Male	No	2	1
12	Bachelors	2012	Bangalore	3	24	Male	No	2	1
13	Bachelors	2012	Bangalore	3	24	Male	No	2	0
14	Bachelors	2018	Bangalore	3	24	Male	No	2	1
15	PHD	2013	Bangalore	2	25	Male	No	3	1
16	Bachelors	2017	Bangalore	3	25	Female	No	3	0
17	Bachelors	2015	Bangalore	2	25	Female	No	3	1

Results per page: 50 1 – 50 of 1200