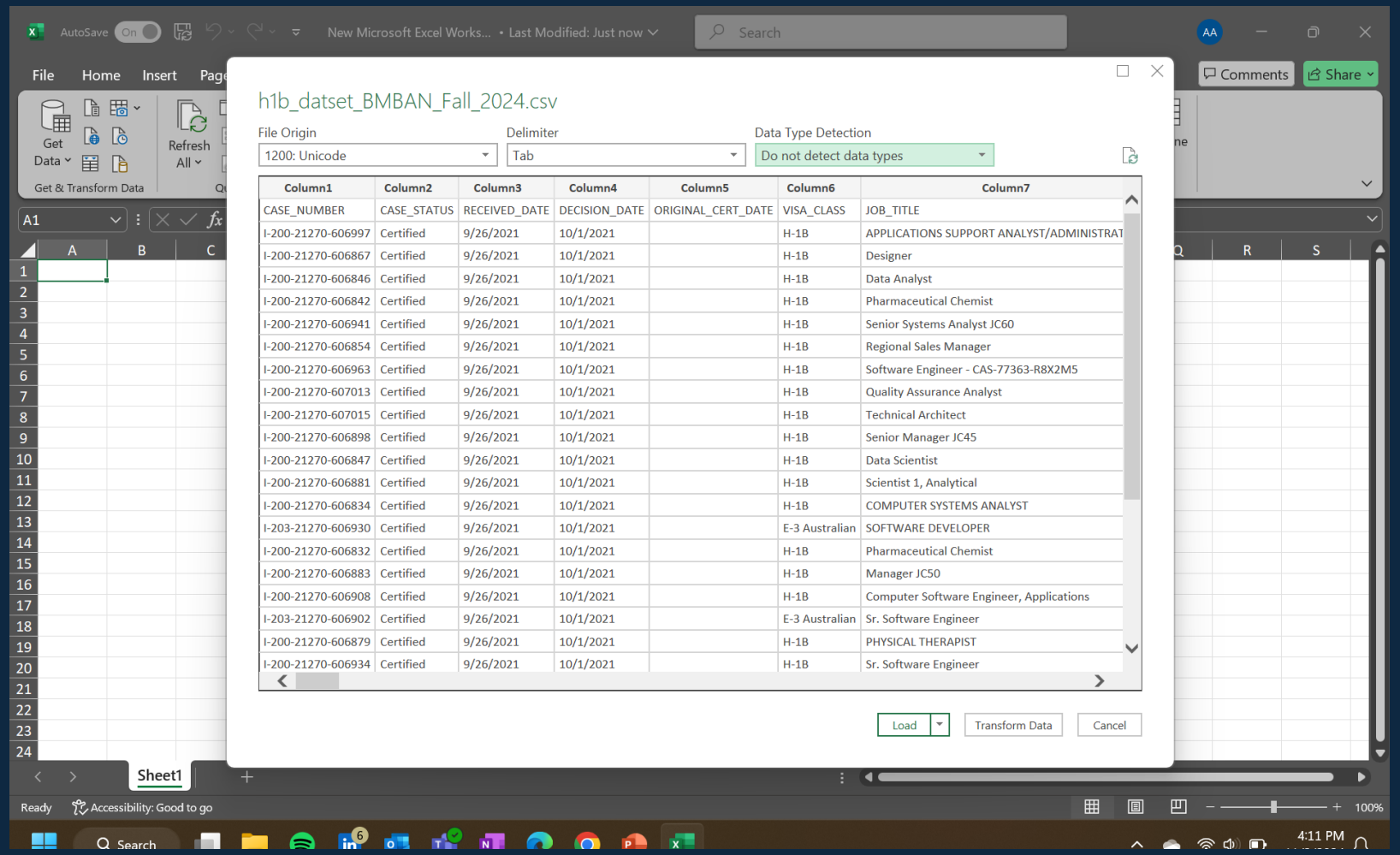
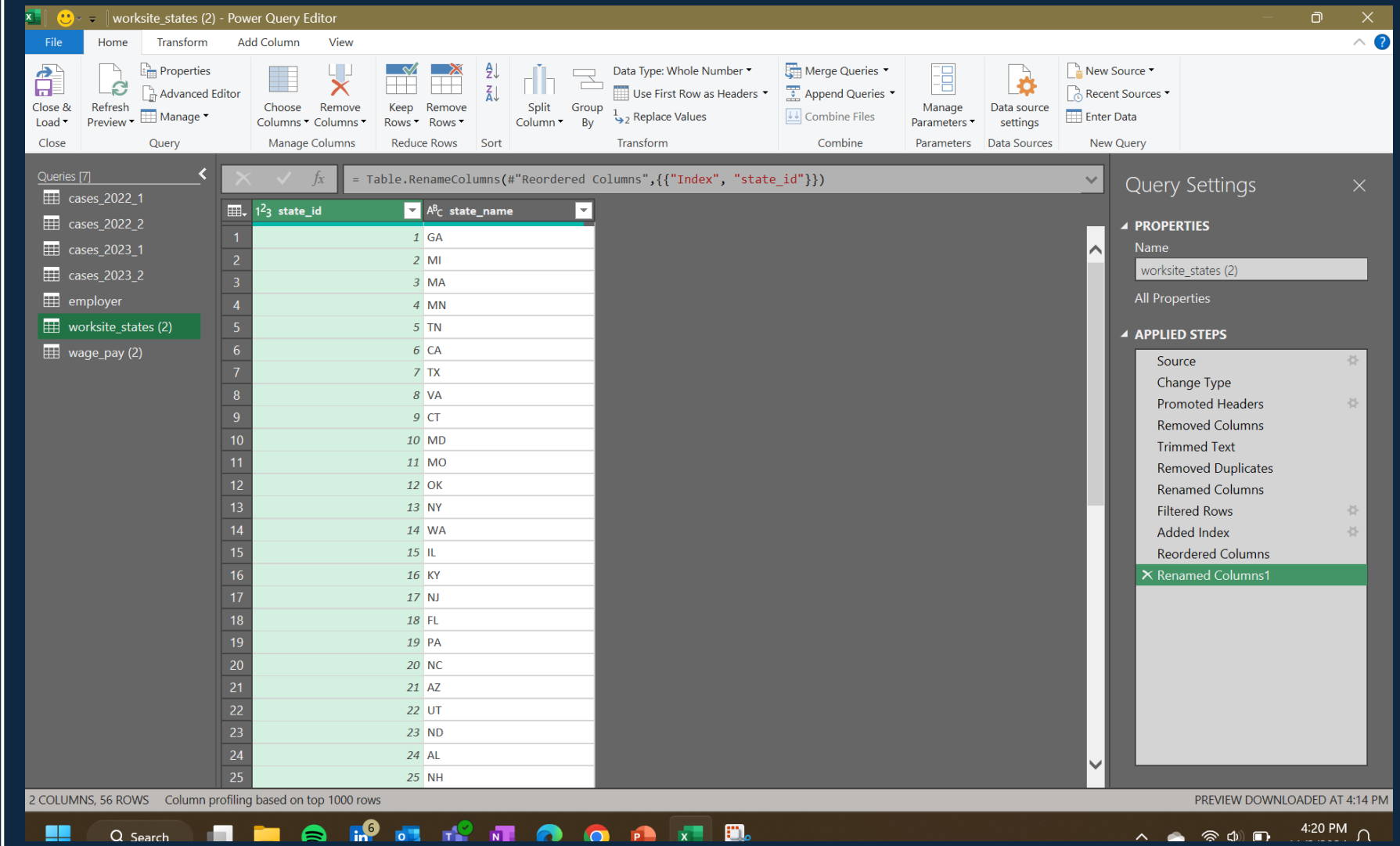
Step – 1: Normalization – Performed the ETL process

Extract



Column1	Column2	Column3	Column4	Column5	Column6	Column7
CASE_NUMBER	CASE_STATUS	RECEIVED_DATE	DECISION_DATE	ORIGINAL_CERT_DATE	VISA_CLASS	JOB_TITLE
I-200-21270-606997	Certified	9/26/2021	10/1/2021		H-1B	APPLICATIONS SUPPORT ANALYST/ADMINISTRAT
I-200-21270-606867	Certified	9/26/2021	10/1/2021		H-1B	Designer
I-200-21270-606846	Certified	9/26/2021	10/1/2021		H-1B	Data Analyst
I-200-21270-606842	Certified	9/26/2021	10/1/2021		H-1B	Pharmaceutical Chemist
I-200-21270-606941	Certified	9/26/2021	10/1/2021		H-1B	Senior Systems Analyst JC60
I-200-21270-606854	Certified	9/26/2021	10/1/2021		H-1B	Regional Sales Manager
I-200-21270-606963	Certified	9/26/2021	10/1/2021		H-1B	Software Engineer - CAS-77363-R8X2M5
I-200-21270-607013	Certified	9/26/2021	10/1/2021		H-1B	Quality Assurance Analyst
I-200-21270-607015	Certified	9/26/2021	10/1/2021		H-1B	Technical Architect
I-200-21270-606898	Certified	9/26/2021	10/1/2021		H-1B	Senior Manager JC45
I-200-21270-606847	Certified	9/26/2021	10/1/2021		H-1B	Data Scientist
I-200-21270-606881	Certified	9/26/2021	10/1/2021		H-1B	Scientist 1, Analytical
I-200-21270-606834	Certified	9/26/2021	10/1/2021		H-1B	COMPUTER SYSTEMS ANALYST
I-203-21270-606930	Certified	9/26/2021	10/1/2021		E-3 Australian	SOFTWARE DEVELOPER
I-200-21270-606832	Certified	9/26/2021	10/1/2021		H-1B	Pharmaceutical Chemist
I-200-21270-606883	Certified	9/26/2021	10/1/2021		H-1B	Manager JC50
I-200-21270-606908	Certified	9/26/2021	10/1/2021		H-1B	Computer Software Engineer, Applications
I-203-21270-606902	Certified	9/26/2021	10/1/2021		E-3 Australian	Sr. Software Engineer
I-200-21270-606879	Certified	9/26/2021	10/1/2021		H-1B	PHYSICAL THERAPIST
I-200-21270-606934	Certified	9/26/2021	10/1/2021		H-1B	Sr. Software Engineer

Step – 1: Normalization – Performed the ETL process

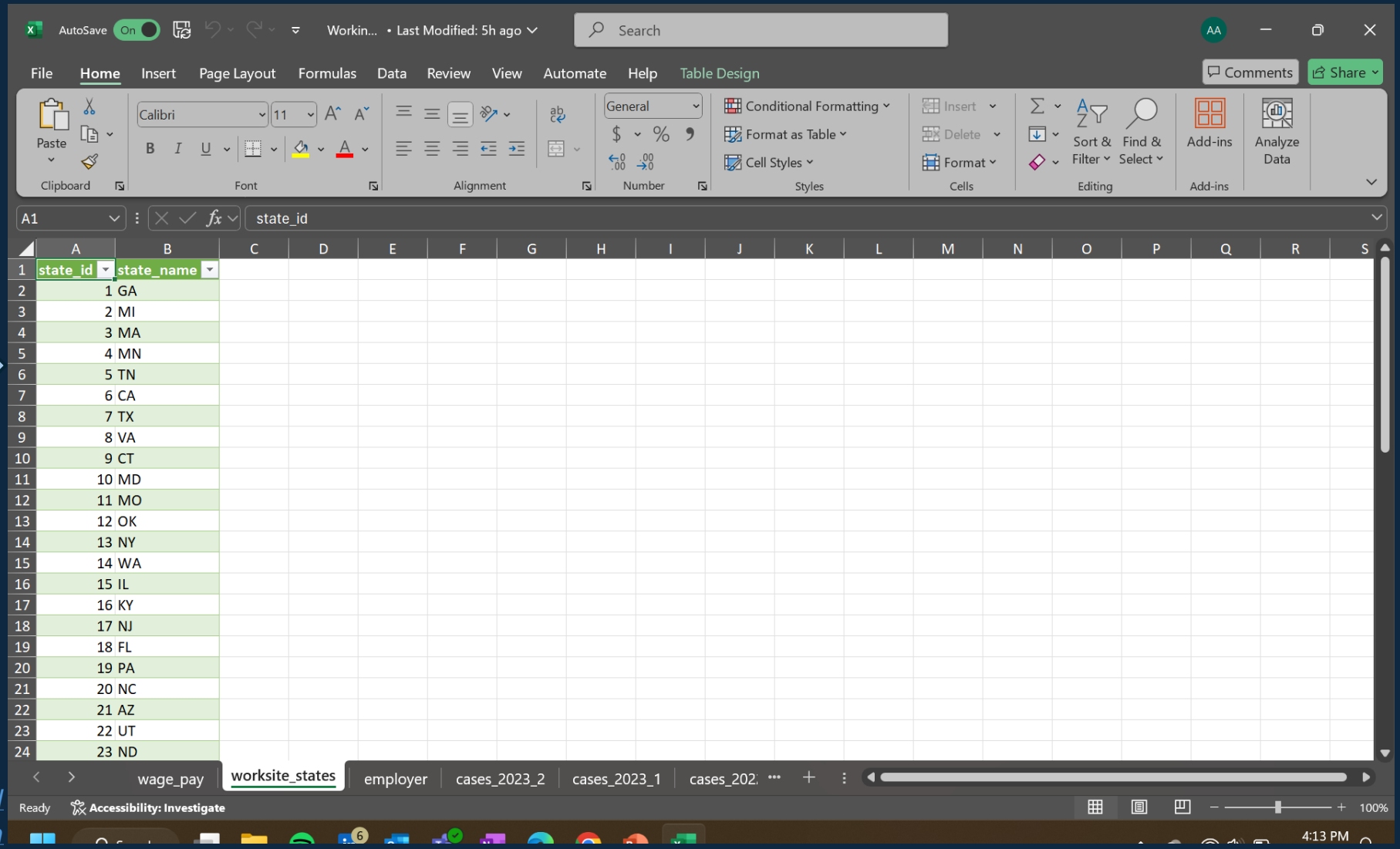


Transform

*** refer the last slide to understand the table distribution*

Step – 1: Normalization – Performed the ETL process

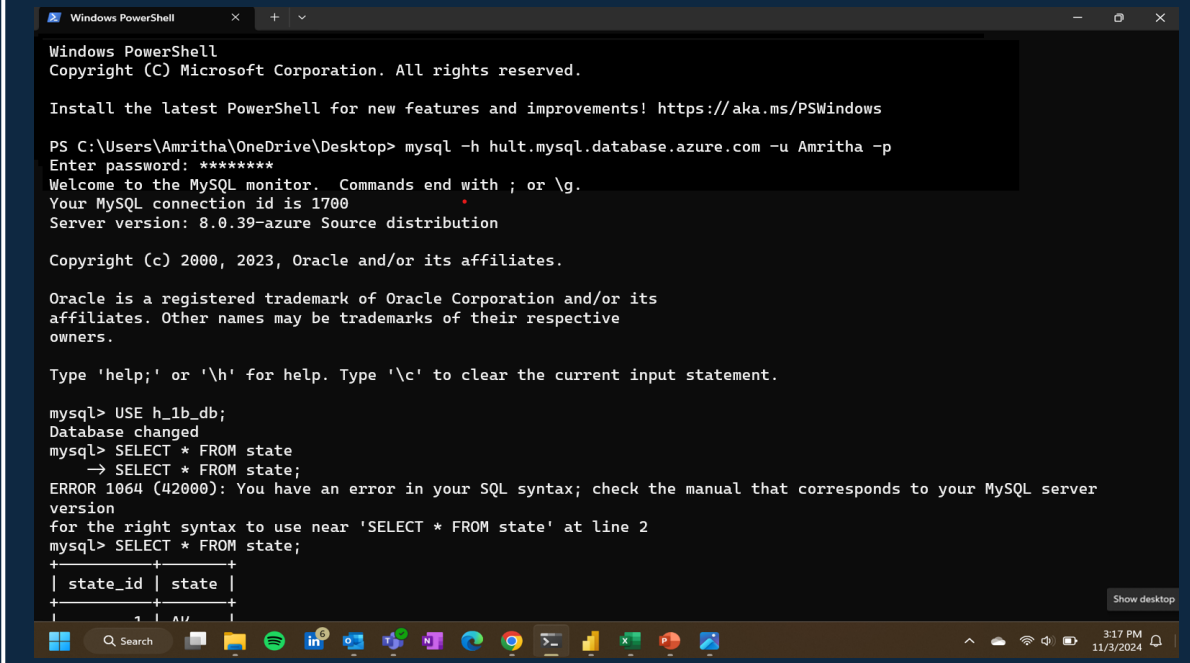
Load



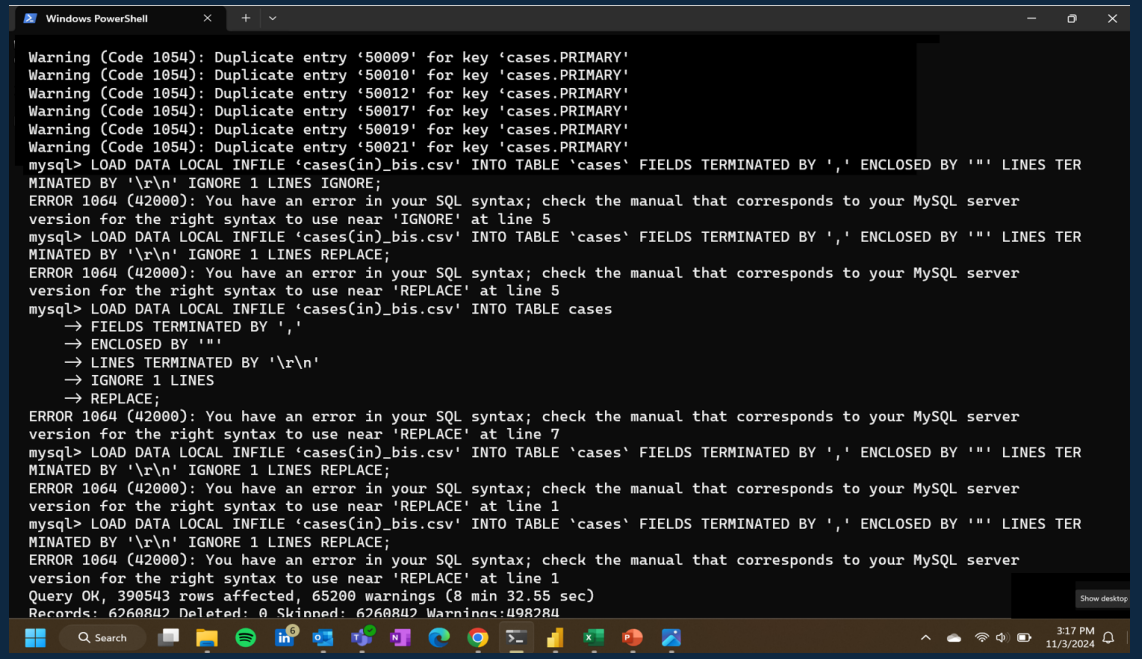
state_id	state_name
1	GA
2	MI
3	MA
4	MN
5	TN
6	CA
7	TX
8	VA
9	CT
10	MD
11	MO
12	OK
13	NY
14	WA
15	IL
16	KY
17	NJ
18	FL
19	PA
20	NC
21	AZ
22	UT
23	ND

*** refer the last slide to understand the table distribution*

Step – 2: Data Uploading – Using Command Prompt

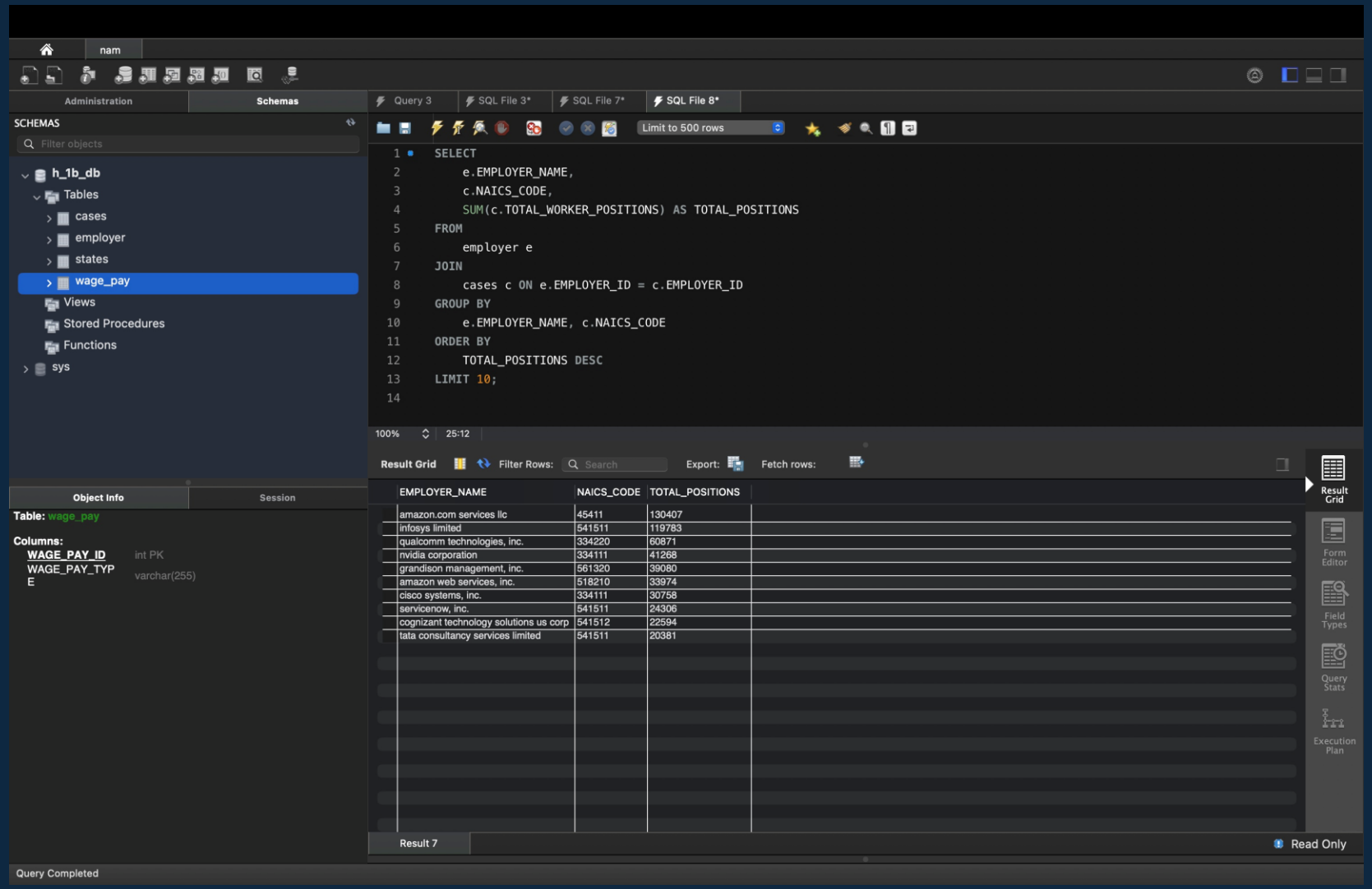


i. Converted the loaded tabled in .csv file type



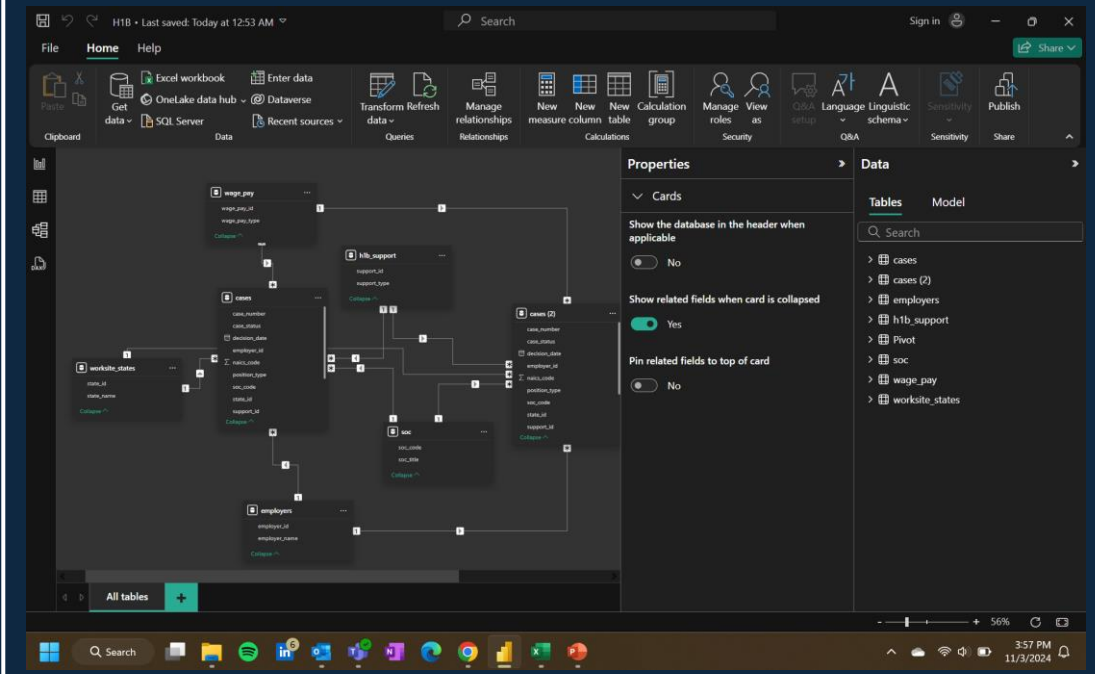
ii. Uploading the data into command prompt

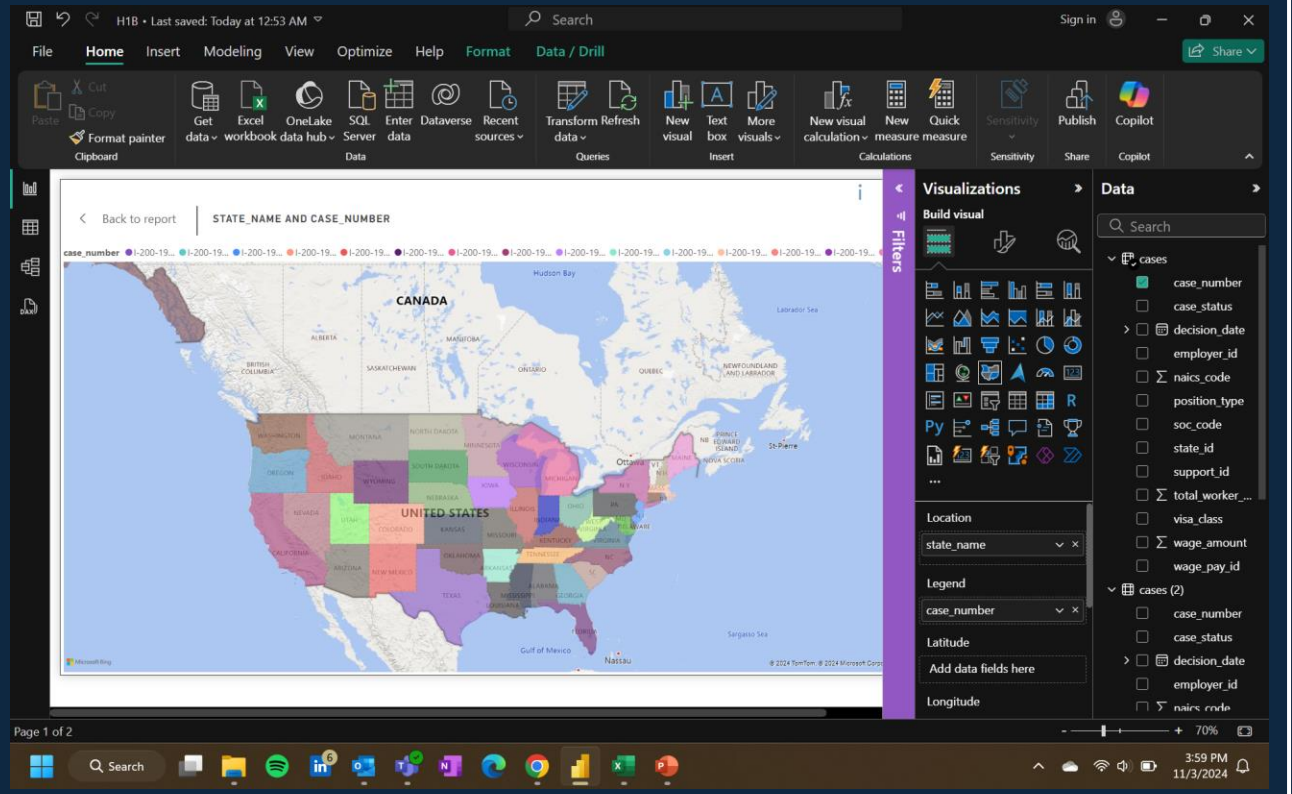
Step – 3: Query Execution – Using MySQL



Step – 4: PowerBI - Visualization

i. Uploaded and merged queries into PowerBI model





ii. Analyzed and created graphs using PowerBI visualization tools

*** refer the last slide to understand the table distribution*

Notes

In the dataset - Total number of records = 1,655,836 & Total number of columns = 96

In 2022,
Number of records from (Q1 to Q4) = 626,084
Number of records (Q1 to Q3) = 507,439

In 2024,
Number of records (Q1 to Q3) = 1,029,752
Out of which 587,378 records are NULL

- Hence, in order to perform fair analysis, we have considered **YTD (Year-to-Date)** data from both the years.
- Considering only Q1, Q2 and Q3 data from the respective years 2022 and 2024.

**

- It can be observed in the previous images that the cases table is mentioned as cases 2022_1, cases 2022_2, cases 2023_1 and cases 2023_2.
- This is because of the huge data records; we have divided the table into two equal halves and worked on it to avoid crashing the laptop.