SQL Query Documentation

Project: CDACL-001-Automobile Insurance Complaint Analysis

Team ID: PTID-CDA-NOV-25-826

Purpose

This document provides runnable SQL queries used for analysis of the automobile insurance complaints dataset.

Each Query includes its purpose, SQL logic, and expected output.

Environment

Database: project_automobile_insurance

SQL Dialect: MySQL-compatible (uses sum(), count(), GROUP BY CASE, ORDER BY)

Permissions: Queries assume SELECT access only (no UPDATE/DELETE operations used).

Inferred Schema (Columns Referenced)

Table: auto_table

- o NAIC
- o Company_Name
- o Ratio
- o Upheld_complaints
- o Questions_of_Fact_complaints
- o Not_upheld_complaints
- o Total_Complaints
- o Premiums_written
- o Ranking
- o Filing_year

High-Level Summary of File Contents

The SQL file contains data validation, aggregation, trend analysis, and insight-driven queries focused on:

- o Dataset inspection
- o Company and year coverage
- o Complaint distribution evaluation
- o Ratio-based performance evaluation
- o Identification of high-risk companies
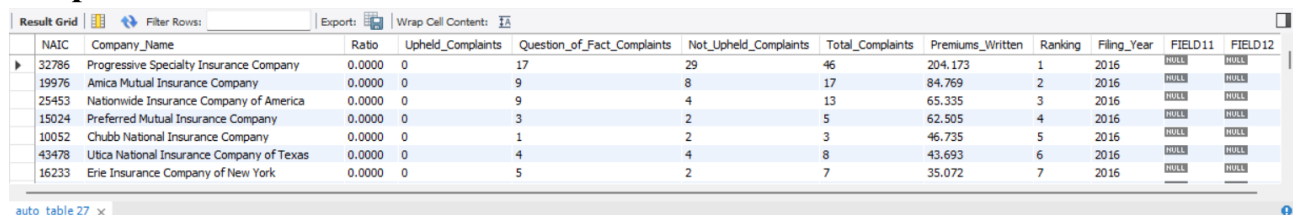
The results are later consumed in Power BI Dashboards

Query Reference – Purpose & Output

Query 0 – Verify table and sample records

Purpose: Inspect table availability and sample data before running analytical queries.
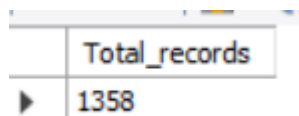
Query: select * from auto_table;

Output:

NAIC	Company_Name	Ratio	Upheld_Complaints	Question_of_Fact_Complaints	Not_Upheld_Complaints	Total_Complaints	Premiums_Written	Ranking	Filing_Year	FIELD11	FIELD12
32786	Progressive Specialty Insurance Company	0.0000	0	17	29	46	204.173	1	2016	NULL	NULL
19976	Amica Mutual Insurance Company	0.0000	0	9	8	17	84.769	2	2016	NULL	NULL
25453	Nationwide Insurance Company of America	0.0000	0	9	4	13	65.335	3	2016	NULL	NULL
15024	Preferred Mutual Insurance Company	0.0000	0	3	2	5	62.505	4	2016	NULL	NULL
10052	Chubb National Insurance Company	0.0000	0	1	2	3	46.735	5	2016	NULL	NULL
43478	Utica National Insurance Company of Texas	0.0000	0	4	4	8	43.693	6	2016	NULL	NULL
16233	Erie Insurance Company of New York	0.0000	0	5	2	7	35.072	7	2016	NULL	NULL

auto_table 27 ×

Query 1 – Total number of Records

Purpose: Determine the total number of rows present in the dataset

Query: select count(*) as Total_records from auto_table;

Total_records
1358

Output:

Query 2 – Filing year coverage

Purpose: Identify all distinct filing years covered in the dataset.

Query: select distinct filing_year from auto_table order by filing_year;

Output:

	filing_year
▶	2009
	2010
	2011
	2012
	2013
	2014
	2015
	2016

Query 3 – Total insurance companies

Purpose: Count the number of distinct insurance companies

Query: select count(distinct Company_Name) as Total_companies from auto_table;

Output:

	Total_companies
▶	223

Query 4 – Null value validation

Purpose: Identify records with missing values in the critical analytical columns.

Query: select

   sum(case when company_name is null then 1 else 0 end) as null_company,

   sum(case when filing_year is null then 1 else 0 end) as null_year,

   sum(case when premiums_written is null then 1 else 0 end) as null_premiums

from auto_table;

Output:

	null_company	null_year	null_premiums
▶	0	0	0

Query 5 – Complaint volume by filing year

Purpose: To calculate and compare the total number of complaints filed in each year to identify year-wise trends.

Query: select filing_year,

sum(total_complaints) as total_complaints

from auto_table group by filing_year order by filing_year;

Output:

filing_year	total_complaints
2009	6808
2010	6426
2011	4700
2012	4023
2013	3389
2014	4487
2015	3043
2016	3043

Query 6 – Complaint Type Distribution by filing year

Purpose: To analyze how different complaint outcomes (upheld, question of fact, and not upheld) are distributed across each filing year.

Query: select filing_year,

sum(upheld_complaints) as upheld,

sum(question_of_fact_complaints) as question_of_fact,

sum(not_upheld_complaints) as not_upheld

from auto_table group by filing_year order by filing_year;

Output:

filing_year	upheld	question_of_fact	not_upheld
2009	966	2104	3738
2010	621	2395	3410
2011	478	1423	2799
2012	459	1190	2374
2013	520	972	1897
2014	610	1382	2495
2015	281	1004	1758
2016	281	1004	1758

Query 7 – Complaint Type Percentage Contribution by Filing Year

Purpose: To calculate the percentage share of each complaint outcomes type relative to total complaints for each filing year.

Query: select filing_year,

round(sum(upheld_complaints) *100.0 / sum(total_complaints),2) as pct_upheld,

round(sum(question_of_fact_complaints) *100.0 / sum(total_complaints),2) as pct_qof,

round(sum(not_upheld_complaints) *100.0 / sum(total_complaints),2) as pct_not_upheld

from auto_table group by filing_year order by filing_year;

Output:

filing_year	pct_upheld	pct_qof	pct_not_upheld
2009	14.19	30.90	54.91
2010	9.66	37.27	53.07
2011	10.17	30.28	59.55
2012	11.41	29.58	59.01
2013	15.34	28.68	55.98
2014	13.59	30.80	55.61
2015	9.23	32.99	57.77
2016	9.23	32.99	57.77

Query 8 – Total Complaints by Company

Purpose: To identify and compare insurance companies based on the total number of complaints received, ranked from highest to lowest.

Query: select company_name,

sum(total_complaints) as Total_complaints

from auto_table group by company_name order by Total_complaints desc;

company_name	Total_complaints
Allstate Insurance Company	4949
GEICO General Insurance Company	3898
Country-Wide Insurance Company	3086
State Farm Mutual Automobile Insurance Comp...	2629
GEICO Indemnity Company	2057
Government Employees Insurance Company	1797
Liberty Mutual Fire Insurance Company	1113
Allstate Property and Casualty Insurance Comp...	923

Output:

Query 9 – Upheld Complaints by Company

Purpose: To compare insurance companies based on the total number of upheld complaints, ranked from highest to lowest.

Query: select company_name,

       sum(upheld_complaints) as Upheld_complaints

from auto_table group by company_name order by Upheld_complaints desc;

company_name	Upheld_complaints
Allstate Insurance Company	760
Country-Wide Insurance Company	368
State Farm Mutual Automobile Insurance Comp...	326
Long Island Insurance Company	265
Allstate Property and Casualty Insurance Comp...	172
GEICO General Insurance Company	171
Liberty Mutual Fire Insurance Company	132
GEICO Indemnity Company	70

Output:

Query 10 – Year-wise Complaints for GEICO Indemnity Company

Purpose: To analyze the annual trend of total complaints filed against GEICO Indemnity Company across different filing years.

Query: select filing_year,

       sum(total_complaints) as Complaints

from auto_table where Company_Name = 'GEICO Indemnity Company'

group by filing_year order by filing_year;

filing_year	Complaints
2009	348
2010	312
2011	189
2012	217
2013	185
2014	286
2015	260
2016	260

Output:

Query 11 – Premium Trend by Filing Year

Purpose: To analyze how total insurance premiums written change across filing years.

Query: select filing_year,

sum(premiums_written) as Total_Premiums

from auto_table group by filing_year order by filing_year ;

filing_year	Total_Premiums
2009	9712.951
2010	10027.525
2011	10283.597
2012	10609.529
2013	10949.204
2014	11290.648
2015	12276.911
2016	12276.911

Output:

Query 12 – Company-wise Premium contribution

Purpose: To compare insurance companies based on their total premium contributions and identify top premium-generating companies.

Query: select company_name,

sum(premiums_written) as Premiums

from auto_table group by Company_Name order by Premiums desc;

Output:

company_name	Premiums
▶ GEICO General Insurance Company	13514.383
State Farm Mutual Automobile Insurance Comp…	9557.517
Allstate Insurance Company	7951.467
GEICO Indemnity Company	6099.016
Government Employees Insurance Company	4281.047
Allstate Property and Casualty Insurance Comp…	3887.041
Liberty Mutual Fire Insurance Company	3405.324
New York Central Mutual Fire Insurance Company	1971.809

Query 13 – Complaints vs Premiums by Company

Purpose: To compare total complaints against total premiums for each company to assess complaint volume relative to business size.

Query: select Company_Name,

　　　sum(total_complaints) as Complaints,

　　sum(premiums_written) as Premiums

from auto_table group by Company_Name;

Company_Name	Complaints	Premiums
▶ 21st Century Advantage Insurance Company	85	120.990
21st Century Centennial Insurance Company	63	98.910
21st Century Indemnity Insurance Company	72	88.876
21st Century Insurance Company	63	3.281
21st Century National Insurance Company	90	118.606
21st Century North America Insurance Company	72	161.430
21st Century Preferred Insurance Company	6	9.738
21st Century Premier Insurance Company	13	41.234

Output:

Query 14 – Complaints per Million dollars of Premium by Company

Purpose: To measure complaint intensity by calculating upheld complaints relative to premiums written for each company.

Query: select Company_Name,

　　　round(sum(upheld_complaints) / sum(premiums_written),4) as complaints_per_million

from auto_table group by company_name order by complaints_per_million desc;

Company_Name	complaints_per_million
Clarendon National Insurance Company	285.7143
OneBeacon Insurance Company	125.0000
Greater New York Mutual Insurance Company	76.9231
Zurich American Insurance Company	37.0370
Infinity Casualty Insurance Company	31.2500
Long Island Insurance Company	25.8865
National Union Fire Insurance Company of Pitts...	19.0476
Sentry Insurance A Mutual Company	14.9254

Output: Result 42 ✕

Query 15 – Complaint Ratio Validation by Company and year

Purpose: To validate the complaint ratio by calculating upheld complaints relative to premiums written for each company and filing year.

Query: select Company_Name,filing_year,

round(sum(upheld_complaints) / sum(premiums_written),4) as Calculated_ratio

from auto_table group by company_name, filing_year;

Company_Name	filing_year	Calculated_ratio
21st Century Advantage Insurance Company	2009	0.2089
21st Century Advantage Insurance Company	2010	0.2239
21st Century Advantage Insurance Company	2011	0.0511
21st Century Advantage Insurance Company	2012	0.0511
21st Century Advantage Insurance Company	2013	0.0000
21st Century Advantage Insurance Company	2014	0.0000
21st Century Advantage Insurance Company	2015	0.1508
21st Century Advantage Insurance Company	2016	0.1508

Output:

Query 16 – Year wise Company Ranking by Complaints Ratio

Purpose: To rank insurance companies within each filing year based on their complaint ratio, where a lower ratio indicates better performance.

Query: select a.Company_Name, a.filing_year, a.ratio,

(

select count(distinct b.ratio) from auto_table b

where b.filing_year = a.filing_year

and b.ratio < a.ratio)

+1 as company_rank

from auto_table a

order by a.filing_year, company_rank;

Output:

	Company_Name	filing_year	ratio	company_rank
	Pacific Indemnity Company	2009	0.0000	1
	Progressive Northern Insurance Company	2009	0.0090	2
	New South Insurance Company	2009	0.0119	3
	GEICO General Insurance Company	2009	0.0122	4
	Government Employees Insurance Company	2009	0.0153	5
	Erie Insurance Company	2009	0.0165	6
	GEICO Indemnity Company	2009	0.0198	7
	Kemper Independence Insurance Company	2009	0.0209	8

Query 17 – Worst-Performing Companies by Average Complaint Ratio

Purpose: To identify insurance companies with the highest average complaint ratios, indicating poorer overall performance.

Query: select company_name,

avg(ratio) as avg_ratio

from auto_table group by company_name order by avg_ratio desc;

	company_name	avg_ratio
▶	New Hampshire Insurance Company	455.36530000
	Clarendon National Insurance Company	280.44590000
	OneBeacon Insurance Company	128.00410000
	Greater New York Mutual Insurance Company	74.53790000
	Zurich American Insurance Company	36.58450000
	Infinity Casualty Insurance Company	31.56860000
	Victoria Fire & Casualty Company	30.82370000
	Long Island Insurance Company	25.88600000

Output: Result 45 ∨

Query 18 – Comparison of Top 5 Companies by Complaint Ratio

Purpose: To compare the best-performing five insurance companies based on complaints, upheld complaints, premiums, and average complaint ratio for dashboard analysis.

Query: select company_name,

    sum(total_complaints) as Total_complaints,

  sum(upheld_complaints) as upheld_complaints,

  sum(premiums_written) as Premiums,

  round(avg(ratio),4) as avg_ratio

from auto_table

group by company_name

order by avg_ratio asc limit 5;

Output:

company_name	Total_complaints	upheld_complaints	Premiums	avg_ratio
Response Worldwide Insurance Company	13	0	25.460	0.0000
Graphic Arts Mutual Insurance Company	3	0	7.293	0.0000
Kemper Financial Indemnity Company	4	0	8.452	0.0000
Vigilant Insurance Company	3	0	34.626	0.0000
National Surety Corporation	2	0	11.535	0.0000

Query 19 – Year with Highest Total Complaints

Purpose: To identify the filing year that recorded the highest number of total complaints.

Query: select filing_year,

    sum(total_complaints) as Complaints

from auto_table group by filing_year order by filing_year desc limit 1;

filing_year	Complaints
2009	6808

Output:

Query 20 – Company Needing Improvement Based on Complaints Ratio

Purpose: To identify the company with the highest average complaint ratio; indicating the greatest need for improvement.

Query: select company_name,

avg(ratio) as Avg_ratio

from auto_table group by company_name order by Avg_ratio desc limit 1;

company_name	Avg_ratio
New Hampshire Insurance Company	455.36530000

Output: