

INSURANCE CLAIMS 2015

SQL queries for insurance claims 2015. We import the file 'insurance_claims.csv' into MySQL Workbench, create a table and modify the variable type and the categories with unusual values. We perform the following queries to extract the information.

1. Number of fraudulent and non-fraudulent claims

What is the total number of fraudulent and non-fraudulent claims?

```
SELECT
SUM(fraud_reported = 'Yes') AS Fraudulent,
SUM(fraud_reported = 'No') AS Non_Fraudulent
FROM claims;
```

Results:

Fraudulent	Non_Fraudulent
247	753

2. Claims by state

How many claims were made in each state?

```
SELECT
incident_state AS incident_state,
count(policy_number) AS claims_by_state
FROM claims
GROUP BY incident_state;
```

Results:

Incident_state	Claims_by_state
NY	262
SC	248
WV	217
NC	110
VA	110
OH	23
PA	30

3. Average annual policy premium by status fraud

What is the average annual policy premium for fraudulent and non-fraudulent claims?

```
SELECT
  fraud_reported AS Fraud,
  ROUND(AVG(policy_annual_premium), 2) AS Avg_Premium
FROM claims
GROUP BY fraud_reported;
```

Results:

Fraud	Avg_Premium
No	1258.43
Yes	1250.24

4. Fraudulent car brand

What is the most common car brand involved in fraudulent claims?

```
SELECT
  auto_make AS Car_Brand,
  COUNT(policy_number) AS Claims
FROM claims
WHERE fraud_reported = 'Yes'
GROUP BY auto_make
ORDER BY Claims DESC LIMIT 1;
```

Results:

Car_Brand	Claims
Ford	22

5. More than two vehicles involved

How many incidents involved more than two vehicles?

```
SELECT
  COUNT(policy_number) AS
  Claims_with_more_than_two_vehicles_involved
FROM claims
WHERE number_of_vehicles_involved > 2;
```

Results:

Claims_with_more_than_two_vehicles_involved
389

6. Top 5 of the most fraudulent cities

Which are the five cities where the most fraud is committed?

```
SELECT
incident_city AS City,
COUNT(policy_number) AS Fraudulent_claims
FROM claims
WHERE fraud_reported = 'Yes'
GROUP BY incident_city
ORDER BY Fraudulent_claims DESC LIMIT 5;
```

Results:

City	Fraudulent_claims
Arlington	44
Columbus	39
Springfield	38
Hillsdale	35
Northbend	34

7. Percentage of fraudulent claims

What percentage of claims were fraudulent?

```
SELECT
ROUND(SUM(fraud_reported = 'Yes') / COUNT(policy_number) * 100, 2)
AS Fraud_Percentage
FROM claims;
```

Results:

Fraud_Percentage
24.70

8. Claims by incident severity

What is the distribution of claims by incident severity?

```
SELECT
incident_severity AS Incident_Severity,
COUNT(policy_number) AS Claims
FROM claims
GROUP BY incident_severity;
```

Results:

Incident_Severity	Claims
Total Loss	280
Minor Damage	354
Major Damage	276
Trivial Damage	90

9. Percentage of claims by type of incident

What percentage of claims were made for each type of incident?

```
SELECT
incident_type AS Incident_Type,
ROUND(100.0 * COUNT(*) / (SELECT COUNT(*) FROM claims), 2) AS Percentage
FROM claims
GROUP BY incident_type
ORDER BY Percentage DESC;
```

Results:

Incident_type	Percentage
Multi-vehicle Collision	41.90
Single Vehicle Collision	40.30
Vehicle Theft	9.40
Parked Car	8.40

10. Percentage of fraudulent claims by type of incident

What percentage of fraudulent claims were made for each type of incident?

```
SELECT
incident_type AS Incident_Type,
ROUND(100.0 * COUNT(*) / (SELECT COUNT(*) FROM claims WHERE fraud_reported = 'Yes'), 2)
AS Percentage
FROM claims
WHERE fraud_reported = 'Yes'
GROUP BY incident_type
ORDER BY Percentage DESC;
```

Results:

Incident_type	Percentage
Single Vehicle Collision	47.37
Multi-vehicle Collision	46.15
Parked Car	3.24
Vehicle Theft	3.24

11. Percentage of claims with witnesses

What percentage of incidents had witnesses present?

```
SELECT  
ROUND(SUM(witnesses > 0) / COUNT(*) * 100, 2) AS Witness_Percentage  
FROM claims;
```

Results:

Witness_Percentage
75.10

12. The top 3 most common brand cars involved in fraudulent claims and the average total claim amount.

What are the top 3 most common brand cars involved in fraudulent claims, and what is their average total claim amount?

```
SELECT  
auto_make AS Auto_Brand,  
COUNT(*) AS Fraudulent_claims,  
ROUND(AVG(total_claim_amount), 2) AS Avg_Claim_Amount  
FROM claims  
WHERE fraud_reported = 'Yes'  
GROUP BY auto_make  
ORDER BY Fraudulent_Claims DESC, Avg_Claim_Amount DESC  
LIMIT 3;
```

Results:

Auto_Brand	Fraudulent_claims	Avg_Claim_Amount
Ford	22	63985
Mercedes	22	53496.82
Chevrolet	21	61873.81

13. The average time between the policy bind date and the incident date.

What is the average time (in months) between the policy start date and the incident date for fraudulent and non-fraudulent claims?

```
SELECT
  fraud_reported AS Fraud,
  ROUND(AVG(TIMESTAMPDIFF(MONTH, policy_bind_date, incident_date)), 2) AS Avg_Months
FROM claims
GROUP BY fraud_reported;
```

Results:

Fraud	Avg_Months
No	155.13
Yes	155.35