ASSIGNMENT

Siri Malatesh Kademani
skademan@asu.edu

Contents

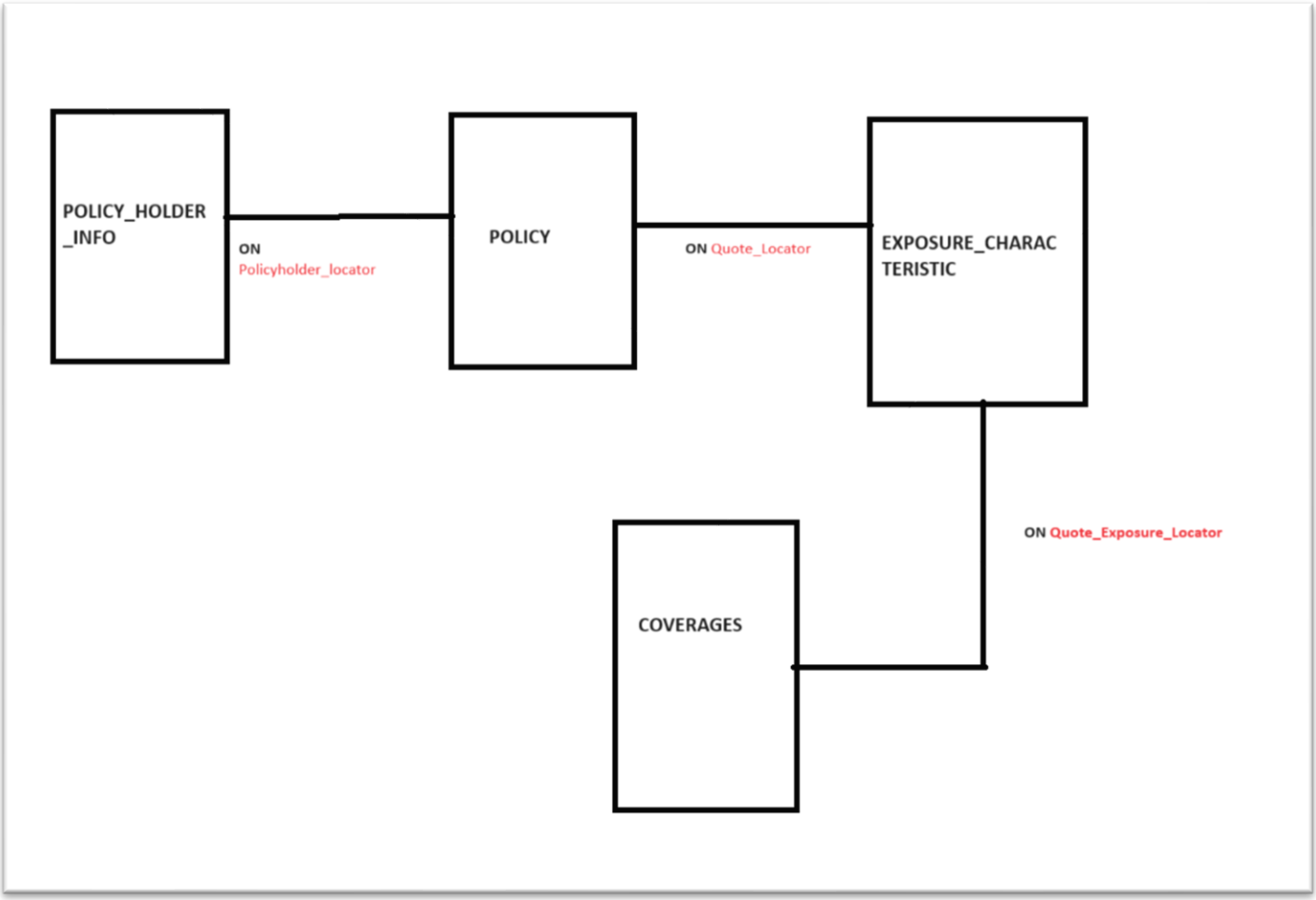
Understanding the Data

Data Glossary

- Prepared a Data Glossary for each sheet.
- The Data Glossary explained each column and how the data was stored in it.
- This helped me in understanding the relationship between POLICYHOLDER, POLICY, EXPOSURE CHARECTERISTICS and COVERAGE relations.

Data_Glossary.xlsx

Relationship Diagram



Data Extract in Snowflake

Step 1: Converted all individual sheets ‘Policy’, ‘Policyholder’, ‘Exposure Characteristics’ and ‘Coverages’ from given input excel file into .csv single files.

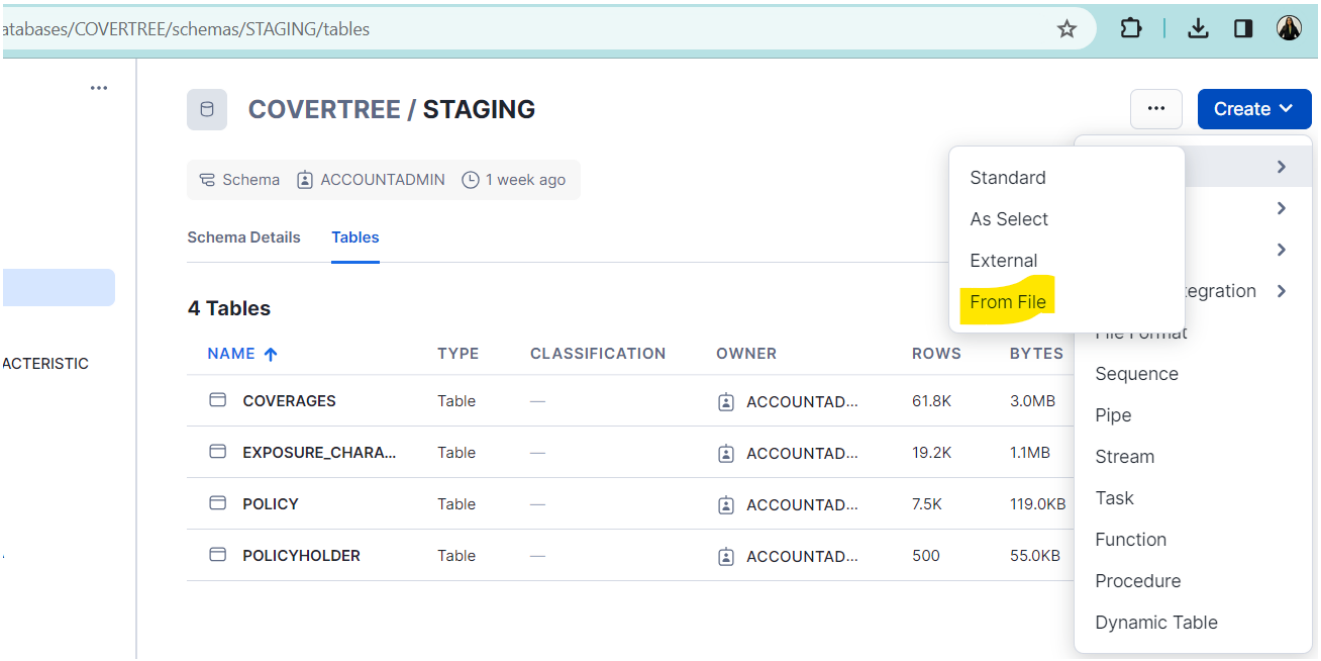
Step 2: Create a database named ‘COVERTREE’.

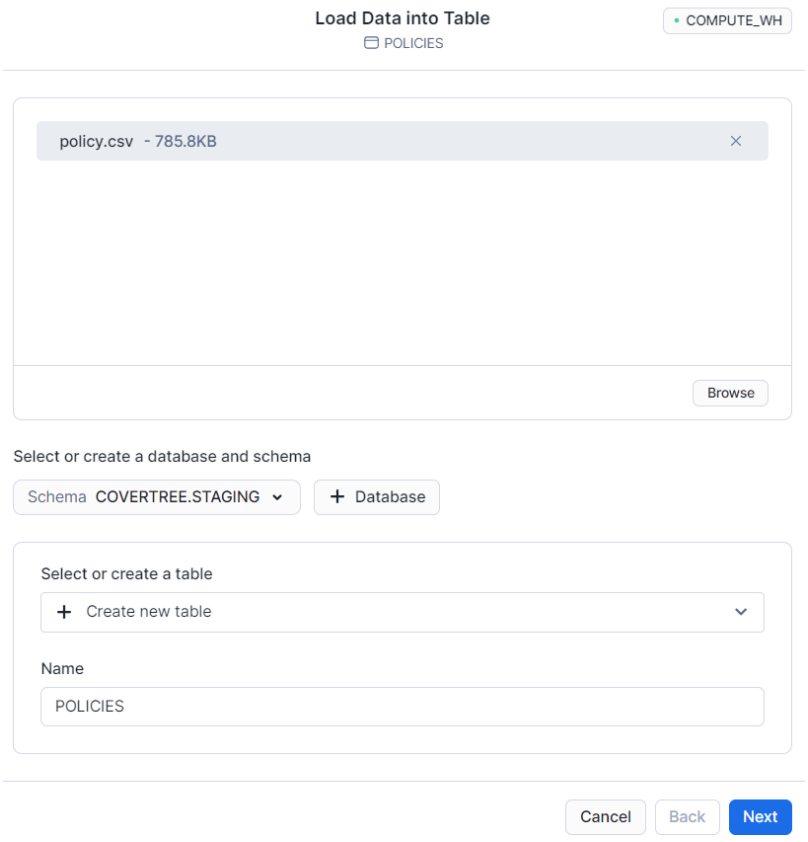
```
CREATE DATABASE COVERTREE;
```

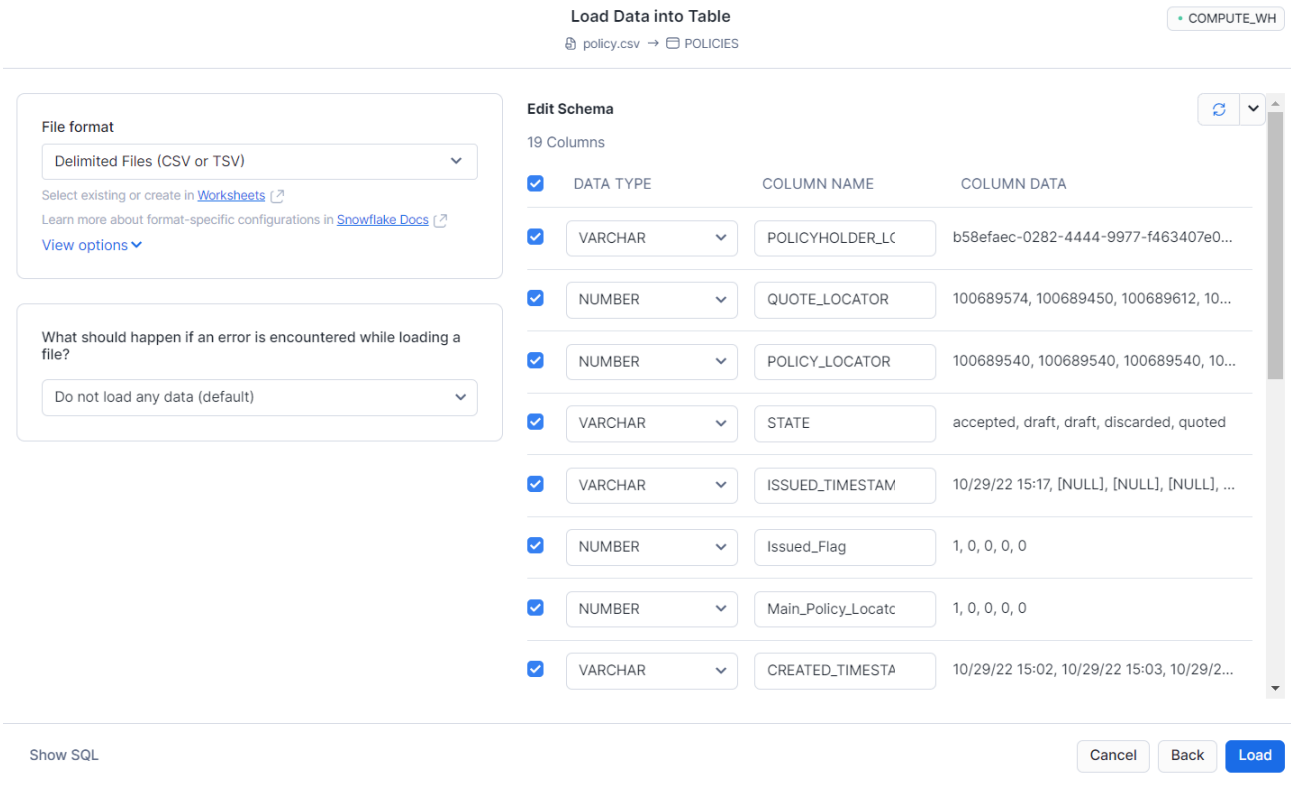
Step 3: Created two schemas ‘STAGING’ and ‘TRANSFORMATIONS’ using Snowflakes SQL worksheet as below:

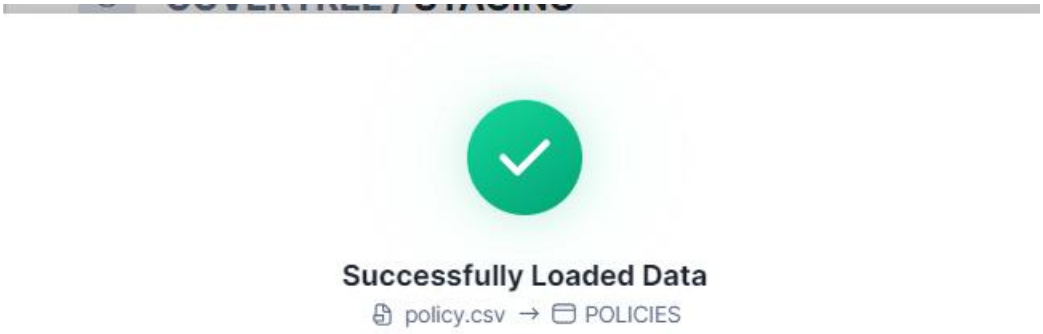
```
CREATE SCHEMA STAGING;

CREATE SCHEMA TRANSFORMATIONS';
```

Step 4: Created staging tables in STAGING schema namely: COVERAGES, EXPOSURE_CHARECTERISTICS, POLICY and POLICYHOLDER.




Load Data into Table
POLICIES
COMPUTE_WH
policy.csv - 785.8KB
Browse
Select or create a database and schema
Schema COVERTREE.STAGING
Database
Select or create a table
Create new table
Name
POLICIES
Cancel
Back
Next


Load Data into Table
policy.csv → POLICIES
COMPUTE_WH
File format
Delimited Files (CSV or TSV)
Select existing or create in Worksheets
Learn more about format-specific configurations in Snowflake Docs
View options
What should happen if an error is encountered while loading a file?
Do not load any data (default)
Edit Schema
19 Columns
DATA TYPE
COLUMN NAME
COLUMN DATA
VARCHAR
POLICYHOLDER_LC
b58efaec-0282-4444-9977-f463407e0...
NUMBER
QUOTE_LOCATOR
100689574, 100689450, 100689612, 10...
NUMBER
POLICY_LOCATOR
100689540, 100689540, 100689540, 10...
VARCHAR
STATE
accepted, draft, draft, discarded, quoted
VARCHAR
ISSUED_TIMESTAM
10/29/22 15:17, [NULL], [NULL], [NULL], ...
NUMBER
Issued_Flag
1, 0, 0, 0, 0
NUMBER
Main_Policy_Locatc
1, 0, 0, 0, 0
VARCHAR
CREATED_TIMESTA
10/29/22 15:02, 10/29/22 15:03, 10/29/2...
Show SQL
Cancel
Back
Load



- Validated the count of rows, data in input file and in snowflake tables after the load was completed.

Data Cleaning and Transformations

Data Cleaning

1. Removed ‘,’ in ‘MODEL_YEAR’ and ‘roof_year_yyyy’ columns.

```
----DATA CLEANING--------------
-- Update the column to store the extracted year value
UPDATE COVERTREE.STAGING.EXPOSURE_CHARACTERISTIC
SET MODEL_YEAR = (REPLACE(MODEL_YEAR, ',', ''));

UPDATE COVERTREE.STAGING.EXPOSURE_CHARACTERISTIC
SET roof_year_yyyy = (REPLACE(roof_year_yyyy, ',', ''));
```

Data Transformations

Tables Created:

POLICY_1 TABLE:

- Replaced String ‘[NULLS]’ as null in POLICY TABLE.
- Converted Date columns which were loaded as Strings into TIMESTAMP datatype.

```
-- CONVERT TEXT DATE VALUES INTO TIME STAMP IN POLICY TABLE:
-- REPLACE STRINGS '[NULL]' as null in POLICY

CREATE OR REPLACE TABLE COVERTREE.TRANSFORMATIONS.POLICY_1 AS (
SELECT
  POLICYHOLDER_LOCATOR,
  QUOTE_LOCATOR,
  POLICY_LOCATOR,
  STATE,
  --ISSUED_TIMESTAMP,
  case when ISSUED_TIMESTAMP='[NULL]' THEN null ELSE TO_TIMESTAMP(ISSUED_TIMESTAMP, 'MM/DD/YY
HH24:MI') END as ISSUED_TIMESTAMP,
  ISSUED_FLAG,
  MAIN_POLICY_LOCATOR_FLAG,
  --CREATED_TIMESTAMP,
```

```
  case when CREATED_TIMESTAMP='[NULL]' THEN null ELSE TO_TIMESTAMP(CREATED_TIMESTAMP,
'MM/DD/YY HH24:MI') END as CREATED_TIMESTAMP,
  --POLICY_START_TIMESTAMP,
  case when POLICY_START_TIMESTAMP='[NULL]' THEN null ELSE
TO_TIMESTAMP(POLICY_START_TIMESTAMP, 'MM/DD/YY HH24:MI') END as POLICY_START_TIMESTAMP,
  --POLICY_END_TIMESTAMP,
  case when POLICY_END_TIMESTAMP='[NULL]' THEN null ELSE TO_TIMESTAMP(POLICY_END_TIMESTAMP,
'MM/DD/YY HH24:MI') END as POLICY_END_TIMESTAMP,
  DATE_OF_BIRTH,
  INSURANCE_SCORE,
  AGENCY_ID,
  AGENT_ID,
  PRIOR_INSURANCE,
  PRIOR_CARRIER_NAME,
  PRIOR_POLICY_EXPIRATION_DATE,
  APPLICATION_INTIATION,
  POLICY_PREMIUM
FROM
  COVERTREE.STAGING.POLICY)
;

select * from  COVERTREE.STAGING.POLICY_1;
```

POLICY_TRANSFORMATION:

Columns created:

- QUOTE_ISSUED_DATE: which contains policy issued date if policy was accepted or if the issued date was missing, then it fills the column with quote created date.
- ISSUED_PREMIUM_AMT: takes premium amount based on whether the policy was accepted.
- POLICY_TENURE_IN_YEARS: tenure of each policy issued.
- AGE: fetches age of policy holder based on DOB.

```
--CREATING QUOTE_ISSUED_DATE:  A DATE THAT CONTAINS ISSUED DATE OF THE POLICY VALUE OR IF THE
ISSUED DATE IS NULL, THEN IT TAKES THE QUOTE CREATED DATE
-- ADDED COLUMN ISSUED_PREMIUM_AMOUNT
-- ADDED COLUMNS POLICY_TENURE_IN_YEARS
-- ADDED AGE COLUMN TO FETCH AGE OF THE POLICY HOLDER
create or replace table COVERTREE.TRANSFORMATIONS.policy_transformation as
(select POLICYHOLDER_LOCATOR,
cast(QUOTE_LOCATOR as varchar(16777216)) as QUOTE_LOCATOR,
POLICY_LOCATOR,
STATE,
ISSUED_TIMESTAMP,
ISSUED_FLAG,
MAIN_POLICY_LOCATOR_FLAG,
CREATED_TIMESTAMP,
(case when issued_timestamp is not null then issued_timestamp else created_timestamp end )as
quote_issued_timestamp,
DATE(case when issued_timestamp is not null then issued_timestamp else created_timestamp end
)as quote_issued_date,
```

```
POLICY_START_TIMESTAMP,
POLICY_END_TIMESTAMP,
ABS(DATEDIFF(MONTH, POLICY_END_TIMESTAMP, POLICY_START_TIMESTAMP) ) AS
POLICY_TENURE_IN_MONTHS,
DATE_OF_BIRTH,
DATE_PART('year', CURRENT_DATE()) - DATE_PART('year', DATE_OF_BIRTH) -
    CASE
        WHEN DATE_PART('month', CURRENT_DATE()) < DATE_PART('month', DATE_OF_BIRTH) THEN 1
        WHEN DATE_PART('month', CURRENT_DATE()) = DATE_PART('month', DATE_OF_BIRTH) AND
             DATE_PART('day', CURRENT_DATE()) < DATE_PART('day', DATE_OF_BIRTH) THEN 1
        ELSE 0
    END AS AGE,
INSURANCE_SCORE,
AGENCY_ID,
AGENT_ID,
PRIOR_INSURANCE,
PRIOR_CARRIER_NAME,
PRIOR_POLICY_EXPIRATION_DATE,
APPLICATION_INTIATION,
POLICY_PREMIUM,
(CASE WHEN STATE= 'accepted' then policy_premium else 0 end )as ISSUED_PREMIUM_AMT
from COVERTREE.TRANSFORMATIONS.policy_1);

SELECT * FROM COVERTREE.TRANSFORMATIONS.policy_transformation;
```

EXPOSURE_CHARECTERISTIC_TRANSFORM:

- Removed columns that contained only/ mostly null values as they hardly add value for further data analysis and insight generation.
- Replaced string '[NULLS]' with nulls and Calculated columns : MODEL_AGE and ROOF_AGE based on model_year and roof_year_yyyy columns.

```
--TRANSFORMATION for EXPOSURE_CHARACTERISTIC table
--remove null columns VACANCY_REASON, FOUR_FEET_FENCE, DIVING_BOARD, VISITORS_IN_A_MONTH,
SKIRTING_TYPE, STORM_MITIGATION_FORTIFIED.
-- Calculate Model age and Roof age of the property
CREATE OR REPLACE TABLE COVERTREE.TRANSFORMATIONS.EXPOSURE_CHARACTERISTIC_TRANSFORM AS(
SELECT
QUOTE_LOCATOR,
QUOTE_EXPOSURE_LOCATOR,
QUOTE_EXPOSURE_CHARACTERISTICS_LOCATOR,
COUNTRY,
STATE,
STREET_ADDRESS,
ZIP_CODE,
CITY,
COUNTY,
LAT,
LONG,
```

```
PROPERTY_WITH_FIRE_PROTECTION,
POLICY_USAGE,
BUSINESS_ON_PREMISES,
FORM,
UNIT_ID,
COMMUNITY_POLICY_DISCOUNT,
UNIT_LOCATION,
PERSONALIZED_PLAN_TYPE,
HOME_TYPE,
ROOF_SHAPE,
SOURCE_OF_HEAT,
ROOF_CONDITION,
MODEL_YEAR,
CASE WHEN MODEL_YEAR='[NULL]' THEN NULL ELSE YEAR(CURRENT_DATE()) - (CAST(MODEL_YEAR AS
INTEGER)) END AS MODEL_AGE,
RCV,
SWIMMING_POOL,
MORTGAGE,
MANUFACTURER_NAME,
TOTAL_SQUARE_FOOTAGE,
UNREPAIRED_DAMAGES,
TRAMPOLINE_LIABILITY,
ROOF_YEAR_YYYY,
CASE WHEN ROOF_YEAR_YYYY='[NULL]' THEN NULL ELSE YEAR(CURRENT_DATE()) - (CAST(ROOF_YEAR_YYYY
AS INTEGER)) END AS ROOF_AGE,
SECURE_RAILS,
ACV,
UTILITY_SERVICES,
THERMO_STATIC_CONTROL,
UNIT_IS_TIED,
HOME_FIXTURES,
ROOF_MATERIAL,
UNUSUAL_RISK,
PARK_NAME,
PURCHASE_DATE,
BURGLAR_ALARM,
STORM_MITIGATION_SHUTTERS,
STORM_MITIGATION_IMPACTGLASS,
--STORM_MITIGATION_FORTIFIED,
WROUGHT_IRON,
SHORT_TERM_RENTAL_SURCHARGE,
DAYCARE_ON_PREMISES,
BUSINESS_EMPLOYEES_ON_PREMISES,
SOURCE_OF_HEAT_INSTALLATION,
TYPE_OF_FUEL,
TRAMPOLINE_SAFETY_NET,
--VACANCY_REASON,
--FOUR_FEET_FENCE,
--DIVING_BOARD,
--VISITORS_IN_A_MONTH,
```

```
--SKIRTING_TYPE
from COVERTREE.STAGING.EXPOSURE_CHARACTERISTIC);
```

COVERAGE_DETAILS:

- Removed PERIL_LOCATOR, NAME, FIELD_NAME, FIELD_VALUE columns and pivoted 'field_name' column to contain 'field_value' column values.

```
-- fetching coverage details per QUOTE_EXPOSURE_LOCATOR
CREATE OR REPLACE TABLE COVERTREE.TRANSFORMATIONS.COVERAGE_DETAILS AS(
select
 QUOTE_EXPOSURE_LOCATOR
--, PERIL_LOCATOR
--, NAME
--, FIELD_NAME
--, FIELD_VALUE
,MAX(case when field_name='medical_payment_to_others_limit_per_person' then field_value end)
as "medical_payment_to_others_limit_per_person"
,MAX(case when field_name='identity_fraud_limit' then field_value end) as
"identity_fraud_limit"
,MAX(case when field_name='cov_c_settlement_option' then field_value end) as
"cov_c_settlement_option"
,MAX(case when field_name='loss_of_use_percentage' then field_value end) as
"loss_of_use_percentage"
,MAX(case when field_name='animal_liability_limit' then field_value end) as
"animal_liability_limit"
,MAX(case when field_name='wind_hail_deductible' then field_value end) as
"wind_hail_deductible"
,MAX(case when field_name='fungi_bacteria_property_limit' then field_value end) as
"fungi_bacteria_property_limit"
,MAX(case when field_name='unscheduled_personal_property_limit' then field_value end) as
"unscheduled_personal_property_limit"
,MAX(case when field_name='debris_removal_limit' then field_value end) as
"debris_removal_limit"
,MAX(case when field_name='equipment_breakdown_limit' then field_value end) as
"equipment_breakdown_limit"
,MAX(case when field_name='water_damage_reduced_limit' then field_value end) as
"water_damage_reduced_limit"
,MAX(case when field_name='all_other_perils_deductible' then field_value end) as
"all_other_perils_deductible"
,MAX(case when field_name='other_structures_limit' then field_value end) as
"other_structures_limit"
,MAX(case when field_name='cov_b_settlement_option' then field_value end) as
"cov_b_settlement_option"
,MAX(case when field_name='manufactured_home_limit' then field_value end) as
"manufactured_home_limit"
,MAX(case when field_name='inflation_guard' then field_value end) as "inflation_guard"
,MAX(case when field_name='loss_assessment_limit' then field_value end) as
"loss_assessment_limit"
,MAX(case when field_name='no_of_golf_carts' then field_value end) as "no_of_golf_carts"
```

```
,MAX(case when field_name='secondary_residence_liability_group' then field_value end) as
"secondary_residence_liability_group"
,MAX(case when field_name='personal_liability' then field_value end) as "personal_liability"
,MAX(case when field_name='damage_to_property_of_others' then field_value end) as
"damage_to_property_of_others"
,MAX(case when field_name='earthquake_deductible' then field_value end) as
"earthquake_deductible"
,MAX(case when field_name='cov_a_settlement_option' then field_value end) as
"cov_a_settlement_option"
,MAX(case when field_name='occasional_vacation_rental' then field_value end) as
"occasional_vacation_rental"
,MAX(case when field_name='enhanced_coverage' then field_value end) as "enhanced_coverage"
,MAX(case when field_name='trip_collision' then field_value end) as "trip_collision"
,MAX(case when field_name='mine_sub_add_living_expense_limit' then field_value end) as
"mine_sub_add_living_expense_limit"
,MAX(case when field_name='residence_burglary_limit' then field_value end) as
"residence_burglary_limit"
,MAX(case when field_name='scheduled_personals' then field_value end) as
"scheduled_personals"
,MAX(case when field_name='water_backup_and_sump_overflow_limit' then field_value end) as
"water_backup_and_sump_overflow_limit"
,MAX(case when field_name='premises_liability_limit' then field_value end) as
"premises_liability_limit"
from COVERTREE.STAGING.COVERAGES
GROUP BY
 QUOTE_EXPOSURE_LOCATOR
--, PERIL_LOCATOR
--, NAME
--, FIELD_NAME
--, FIELD_VALUE
);

select * from COVERTREE.transformations.COVERAGE_DETAILS;
```

SQL Code scripts used in Snowflake:

1.

Staging.sql

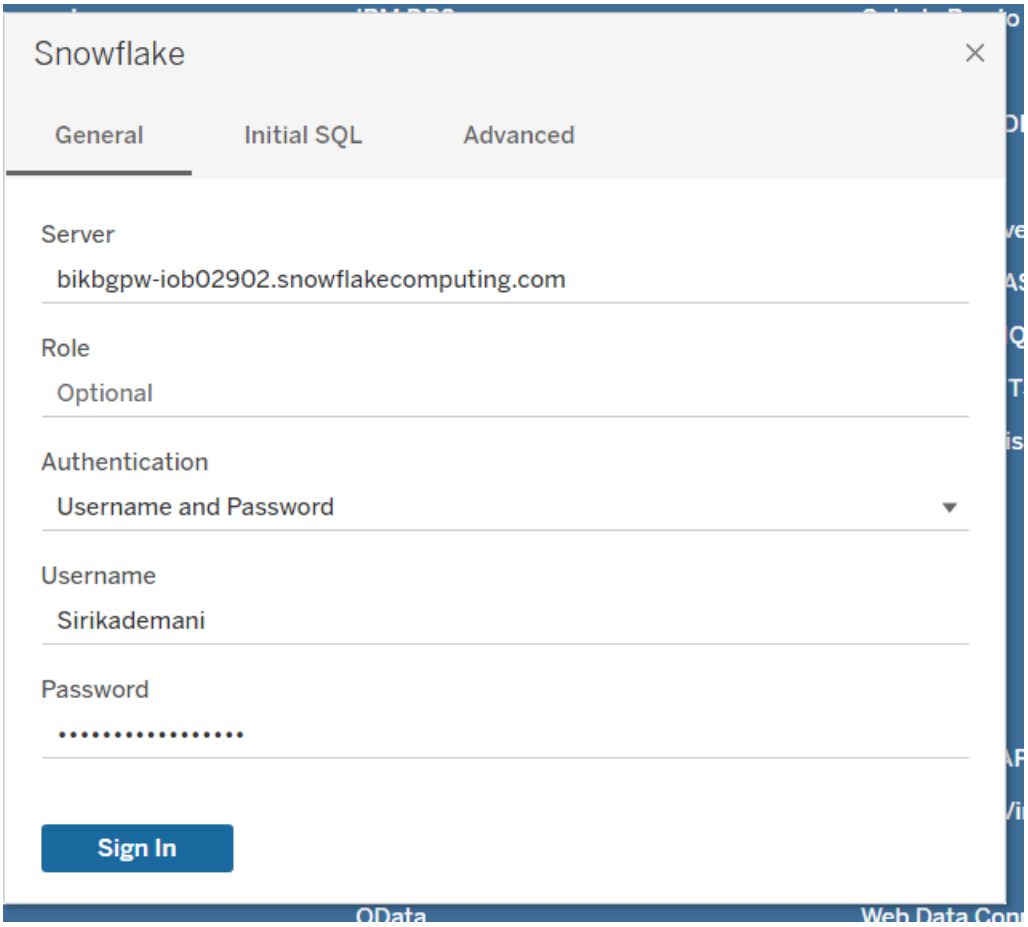
2.

Transformations.sql

Data Extract in Tableau from Snowflake

Connection to Tableau

- Server host: **https://bikbgpw-iob02902.snowflakecomputing.com**
- Warehouse: COMPUTE_WH
- Database: COVERTREE



Data Source Modelling in Tableau.

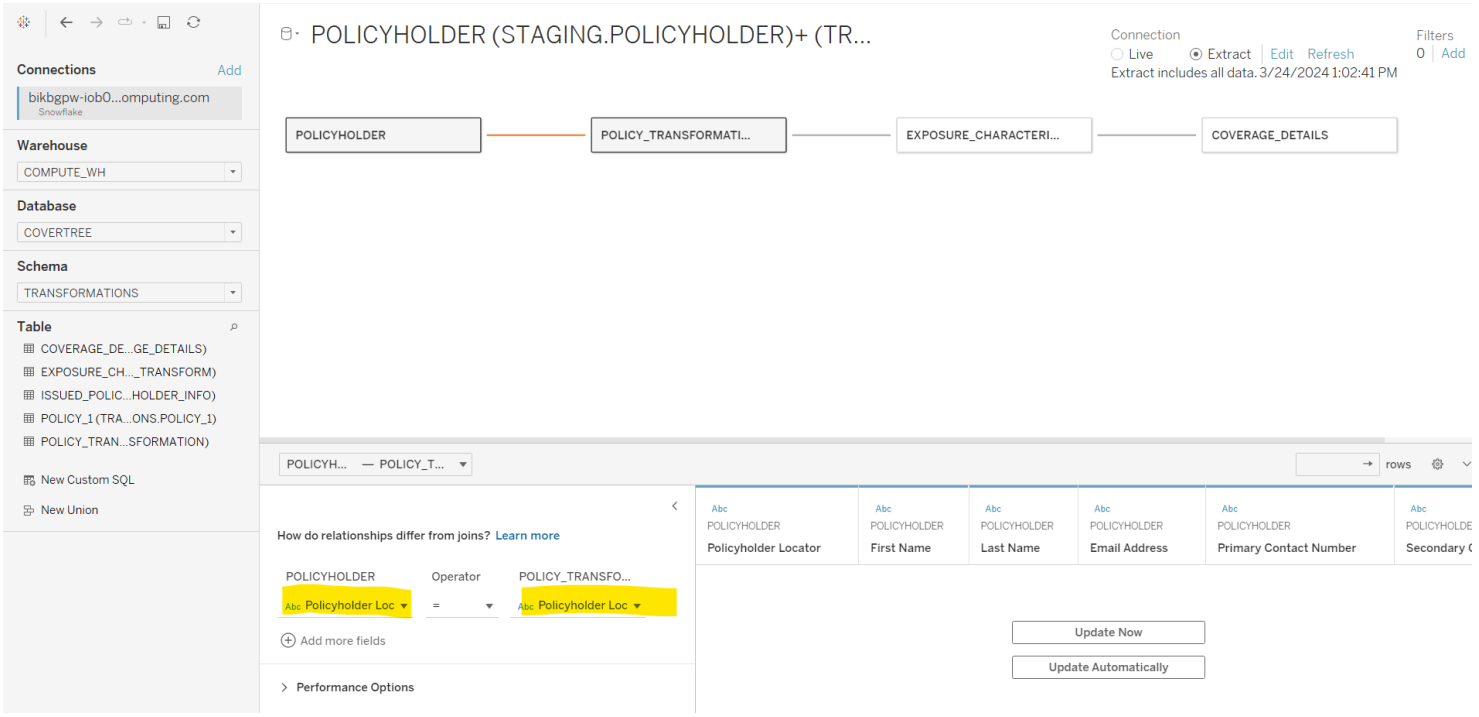
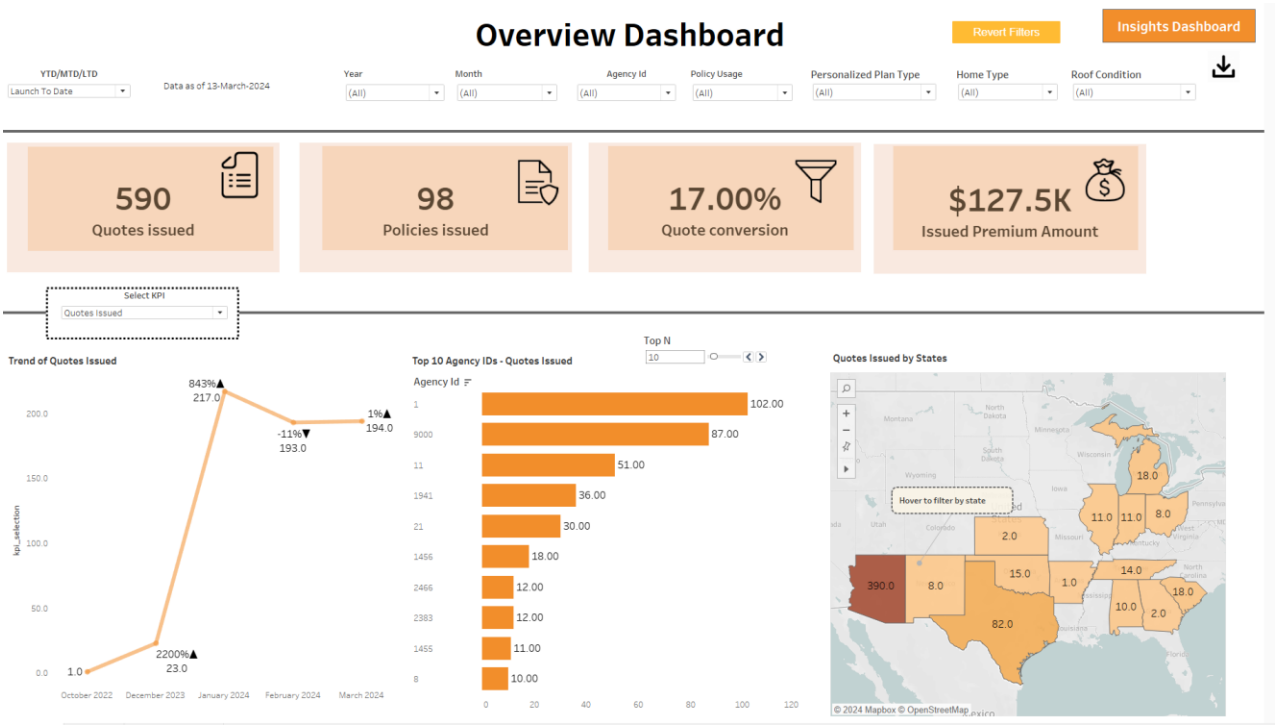
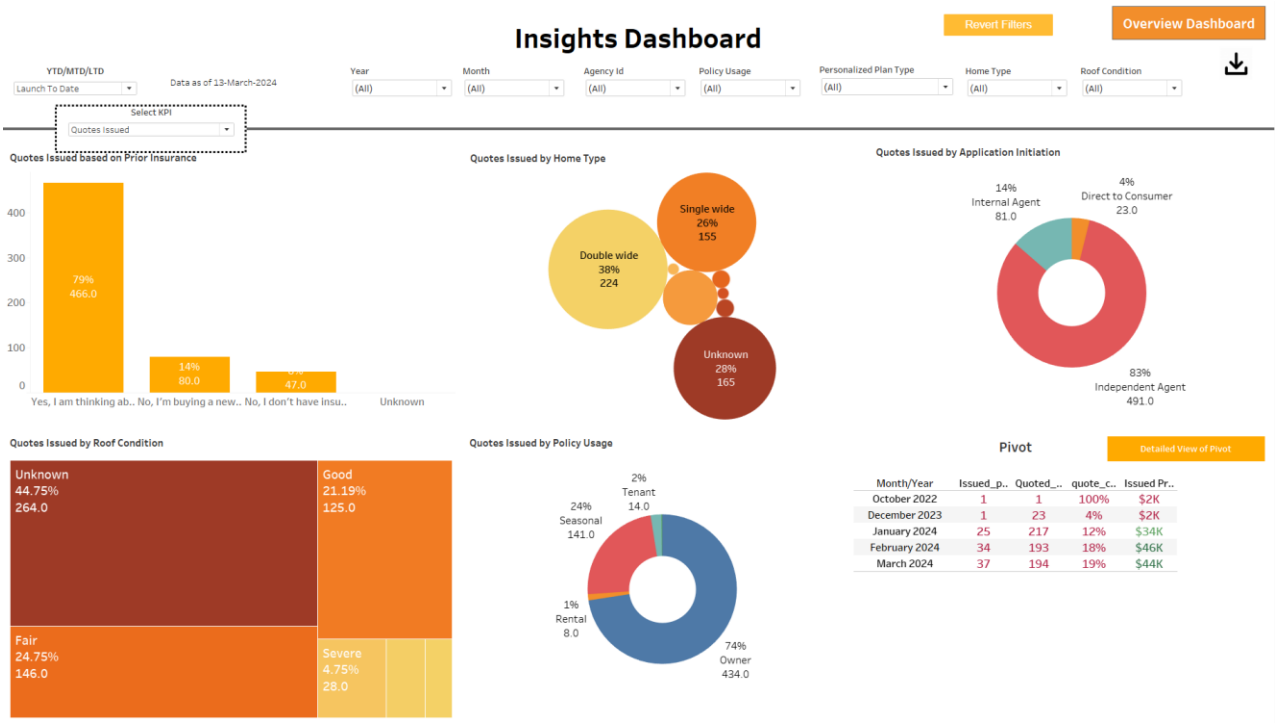


Tableau Dashboards

Overview Dashboard



Insights Dashboard



Pivot Table

Pivot

Insights Dashboard

Month/Year	States	Issued_policies	Quoted_policies	quote_conversion_kpi	Issued Premium Amt
October 2022	AZ	1	1	100%	$2K
December 2023	AZ	1	11	9%	$2K
	IN	0	1	0%	$0K
	MI	0	4	0%	$0K
	NM	0	1	0%	$0K
	SC	0	1	0%	$0K
	TX	0	5	0%	$0K
January 2024	AL	1	6	17%	$2K
	AR	0	1	0%	$0K
	AZ	17	156	11%	$23K
	GA	0	1	0%	$0K
	IL	1	2	50%	$0K
	IN	2	5	40%	$2K
	KS	0	1	0%	$0K
	MI	1	7	14%	$2K
	NM	2	5	40%	$3K
	OH	0	2	0%	$0K
	OK	0	3	0%	$0K
	SC	0	10	0%	$0K
	TN	0	5	0%	$0K
	TX	1	13	8%	$2K
February 2024	AL	0	5	0%	$0K
	AZ	23	105	22%	$31K
	GA	0	1	0%	$0K
	IL	1	7	14%	$2K
	IN	0	3	0%	$0K
	MI	1	4	25%	$1K
	NM	1	2	50%	$2K
	OH	0	4	0%	$0K
	OK	1	11	9%	$2K
	SC	0	4	0%	$0K
	TN	1	8	13%	$0K
	TX	6	39	15%	$7K
March 2024	AZ	26	137	19%	$32K
	IL	0	2	0%	$0K
	IN	0	3	0%	$0K
	KS	1	1	100%	$1K
	MI	0	4	0%	$0K
	NM	0	1	0%	$0K
	OH	1	4	25%	$2K
	OK	2	4	50%	$1K

Tableau Workbook attachment:

CoverTree_Data_Insights.twbx

THANK YOU!