

Data Classification

In this experiment, we will learn to use Snowflake's privacy recommendations function.

1. Create a sample data and table with ID, Social Security number, Age and Credit Card

```
USE ROLE ACCOUNTADMIN;
```

```
CREATE OR REPLACE DATABASE SNOWTEST;  
CREATE OR REPLACE SCHEMA SNOWTEST.DATA_CLASS;  
CREATE OR REPLACE TABLE SNOWTEST.DATA_CLASS.SAMPLE_DATA_TBL(  
    ID VARCHAR(10)  
    , SSN VARCHAR(11)  
    , AGE NUMERIC  
    , CREDIT_CARD VARCHAR(19)  
);
```

2. Let's enter some fake sensitive data

```
INSERT INTO SNOWTEST.DATA_CLASS.SAMPLE_DATA_TBL VALUES  
( 'A0000001', '234-45-6477', 24, '4053-0495-0394-0494' );  
INSERT INTO SNOWTEST.DATA_CLASS.SAMPLE_DATA_TBL VALUES  
( 'A0000002', '234-85-6477', 24, '4653-0495-0394-0494' );  
INSERT INTO SNOWTEST.DATA_CLASS.SAMPLE_DATA_TBL VALUES  
( 'A0000003', '235-45-6477', 24, '4053-0755-0394-0494' );
```

3. Looking at the table values

```
SELECT *  
FROM SNOWTEST.DATA_CLASS.SAMPLE_DATA_TBL;
```

4. Run `EXTRACT_SEMANTIC_CATEGORIES` to analyze the columns in the table

```
SELECT  
EXTRACT_SEMANTIC_CATEGORIES('SNOWTEST.DATA_CLASS.SAMPLE_DATA_TBL');
```

5. The results will be in JSON format. We can flatten to analysis the recommendations

CREATE OR REPLACE TABLE SNOWTEST.DATA_CLASS.CLASSIFICATIONS
AS

```
select t.key::varchar as column_name,  
       t.value:"recommendation".privacy_category::varchar as privacy_category,  
       t.value:"recommendation".semantic_category::varchar as semantic_category,  
       t.value:"recommendation".coverage::numeric as probability  
from table(  
    flatten(  
        extract_semantic_categories(  
            'SNOWTEST.DATA_CLASS.SAMPLE_DATA_TBL'  
        )::variant  
    )  
    ) as t  
;
```

6. We can then use this table to review the recommendation and create object tags accordingly

SELECT * FROM SNOWTEST.DATA_CLASS.CLASSIFICATIONS;

The screenshot shows a Snowflake SQL interface. The top bar includes the user 'ACCOUNTADMIN', the warehouse 'COMPUTE_WH', and a 'Share' button. The main area displays a SQL script with comments and commands to create a table and insert data. Below the script, the 'Results' tab is active, showing a table with 4 rows and 5 columns: COLUMN_NAME, PRIVACY_CATEGORY, SEMANTIC_CATEGORY, and PROBABILITY. The results are as follows:

	COLUMN_NAME	PRIVACY_CATEGORY	SEMANTIC_CATEGORY	PROBABILITY
1	AGE	QUASI_IDENTIFIER	AGE	1
2	CREDIT_CARD	null	null	null
3	ID	null	null	null
4	SSN	IDENTIFIER	NATIONAL_IDENTIFIER	1

On the right side of the results table, there is a 'Query Details' panel showing the query duration as 83ms, 4 rows returned, and the query ID as 01ae9114-0001-4c98-0...

7. Clear Resources

DROP DATABASE SNOWTEST;