# Assignment: Feature Engineering Using Snowflake and Feature Stores

## 1.Introduction to Feature Engineering

### What is Feature Engineering?

Feature Engineering is the process of transforming raw data into meaningful features that enhance the performance of machine learning models. It is one of the most important steps in building predictive models.

### Why is it Important?

Improves model accuracy

Helps in better generalization

Simplifies complex data

### Common Techniques:

**Normalization**: Scaling numeric data (e.g., MinMaxScaler)

**Encoding**: Converting categorical data (e.g., Label Encoding)

**Aggregation**: Time-windowed stats (mean, max, etc.)

**Handling Missing Values**: Dropping, imputing, or replacing nulls

## 2. Using Snowflake for Data Storage & Processing

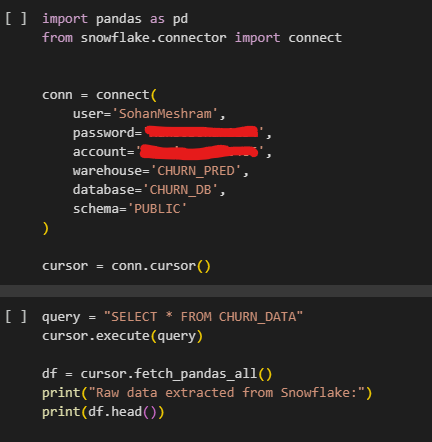
### Snowflake Overview:

Cloud-based data warehouse platform

Supports both structured (tables) and semi-structured data (JSON, XML)

Allows SQL queries for fast analytics and ML preprocessing

### Extracting Data (Code Used):



### Integration with ML Pipelines:

Feature generation in Snowflake → stored in tables

Python SDK (or Snowpark) connects Snowflake to ML pipelines (e.g., scikit-learn, SageMaker, Databricks)

Features fetched into memory for training

## 3. Feature Store Concepts

### What is a Feature Store?

A centralized repository to **store, manage, and serve ML features** consistently for training and inference.

### Why Use a Feature Store?

Centralized and reusable features

Ensures training/serving consistency

Reduces duplication and engineering effort

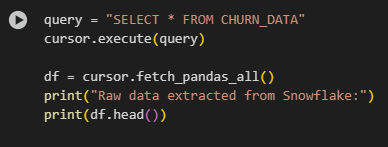
### Comparison Table:

| **Feature Store** | **Strengths** | **Integration** | **Notes** |
| --- | --- | --- | --- |
| Snowflake Feature Store | Native to Snowflake, supports SQL-based transformations | Direct access to Snowflake data | Strong integration with Snowflake |
| AWS SageMaker Feature Store | Managed service, integrates well with AWS ML services | Native AWS services | Good for AWS-centric workflows |
| Databricks Feature Store | Tight integration with Delta Lake and MLflow | Databricks ecosystem | Good for Spark-based workflows |

1. **Implementation with Snowflake & Feature Store**

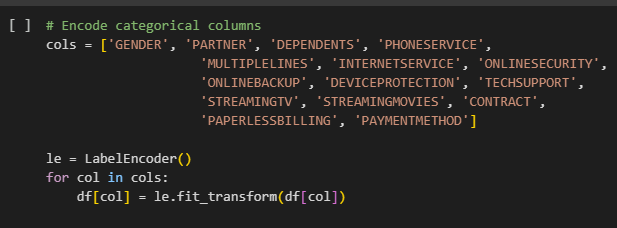
**Extract:**

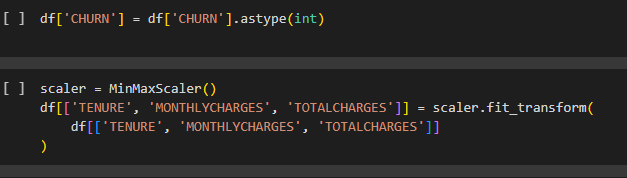
Use Snowflake SQL to extract raw data from tables (as shown in examples)



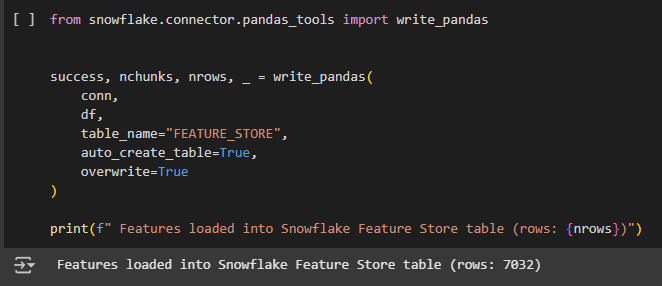
**Transform (Feature Engineering):**

Perform encoding in python, and Normalize using MinMaxScalar.





**Load into Feature Store :**

Create table and Load into featuer store using write\_pandas

**Access for ML Model:**

Extract features from feature store and used for ML

