Sample Model Pipeline Process Flow

Data Sources | **Data Transmission** | **Raw Data** | **Data Preprocessing** | **Model Execution** | **Data Outputs**

Claims DB

Policy DB

1 — *SQL to pull from on-prem DBs* → *Combined Claim and Policy Data*

2 — *Basic Cleaning (type conversions, filters applied, etc.)*

3 — *NLP performed on claims log notes*

NOAA SPC Daily Weather

4 — *Webscraping via Python* → *24 hours of Weather data in 2-minute files*

5 — *Weather Data Joined on Claim/Log Data*

6 — *Weather Data Rolled-up to 1-hour segments*

USFS Tree Canopy

7 — *Webscraping via Python* → *Static Tree Canopy CSV*

8 — *Tree Canopy Data Joined on Weather/Claim/Policy by Coordinates*

9 — *Sample Weather Model* — Calculates loss ration (probability score between 0 an 1)

Loss Ratio + Run Metadata

10 — *Internal Scores DB for Audit*

11 — *Business Partner DB*

Legend

Parquet File		Python		Text — Data Source	
Flat File		Text — Data		Text — Model	