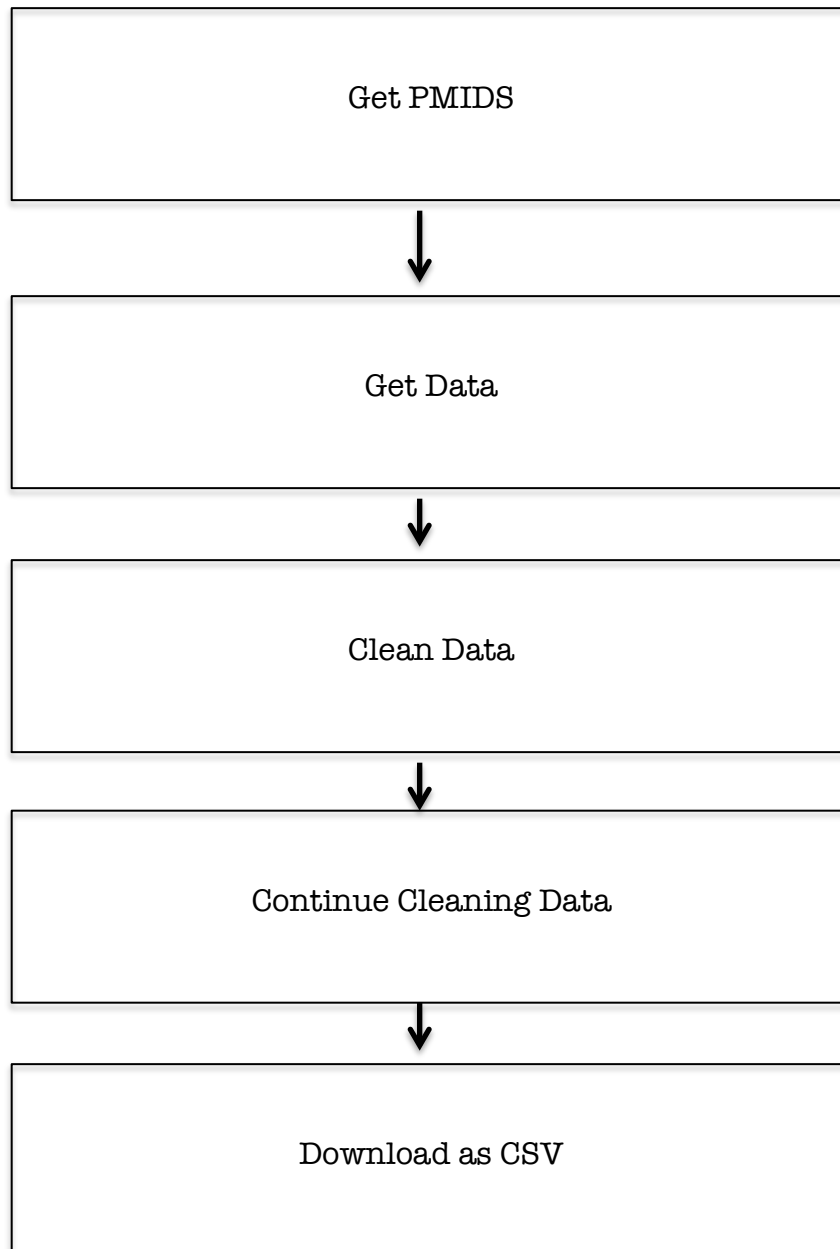


Section 1: Program Design

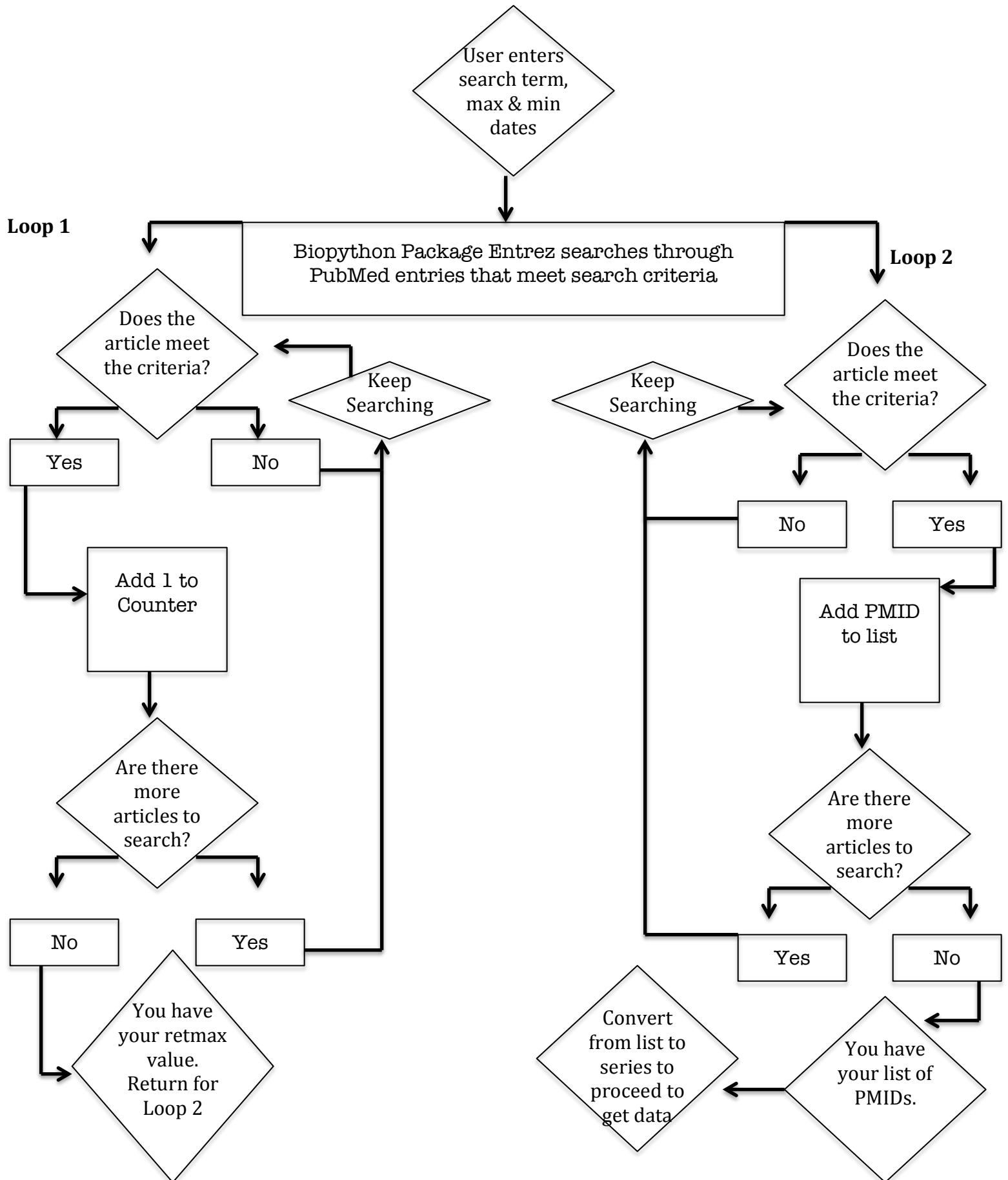
This project is divided into three modules: A data scraper module, a SQL module and a data visualization model. The following pages contain the workflow diagrams detailing the structure and function of these modules.

Data Scraper

These are the five main functions performed in the scraper module, each of which has it's own detailed workflow diagram in the subsequent pages:



Get PMIDs



Get Data

For each PMID in series created in previous step:

Use Entrez efetch function to return record as handle in xml format



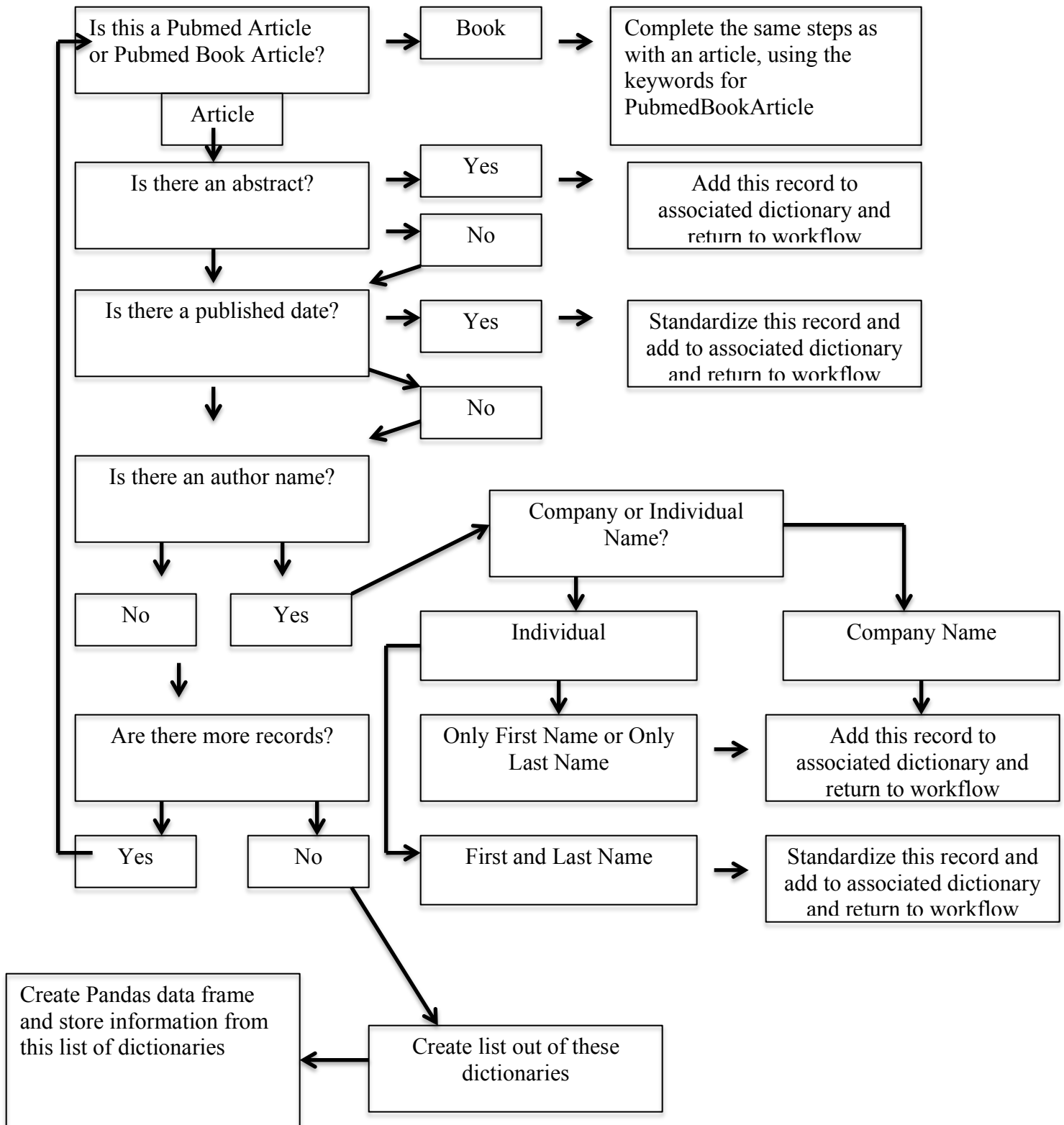
Use Entrez read function to parse XML results into Python dictionary



Append record into dictionary

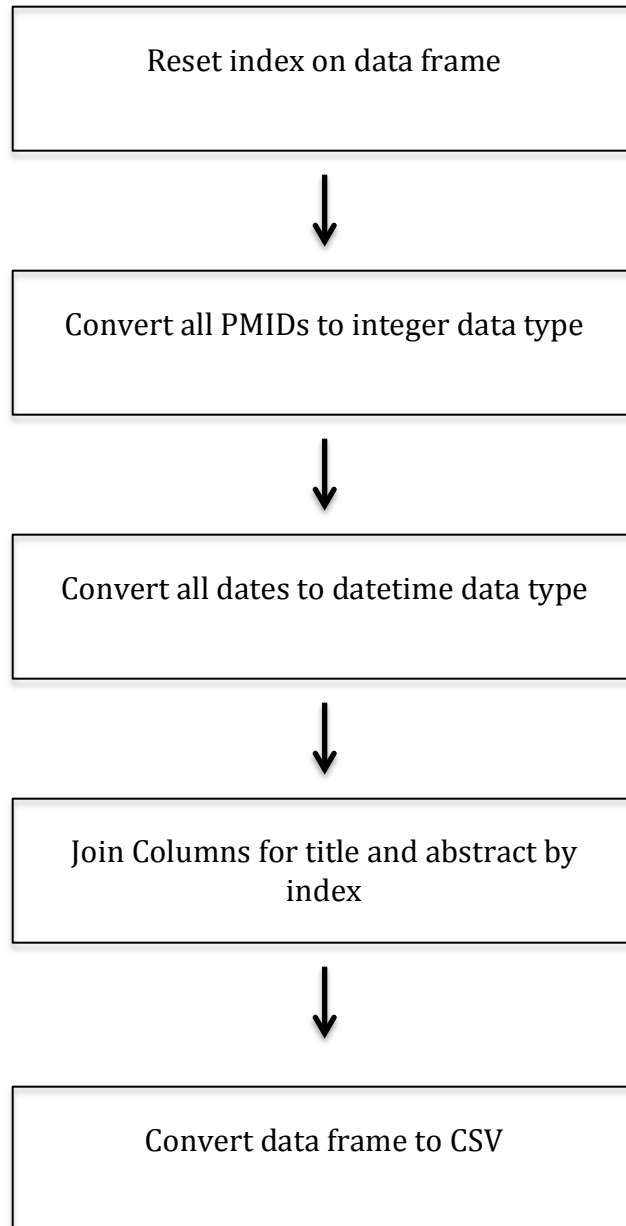
Clean Data

This function uses a list of dictionaries to store all citation data for the dataset. The following is performed for each previously retrieved record:

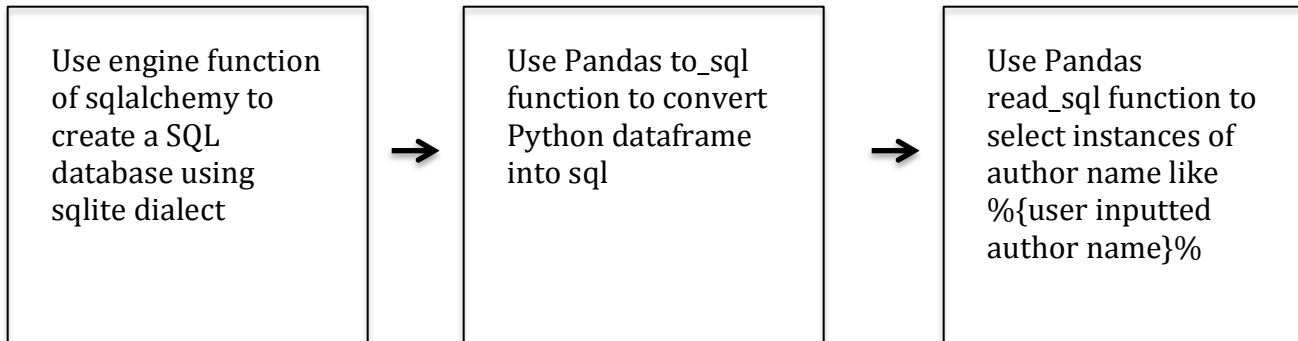


Continue Cleaning Data & Convert to CSV

Begin with dataframe from previous workflow:

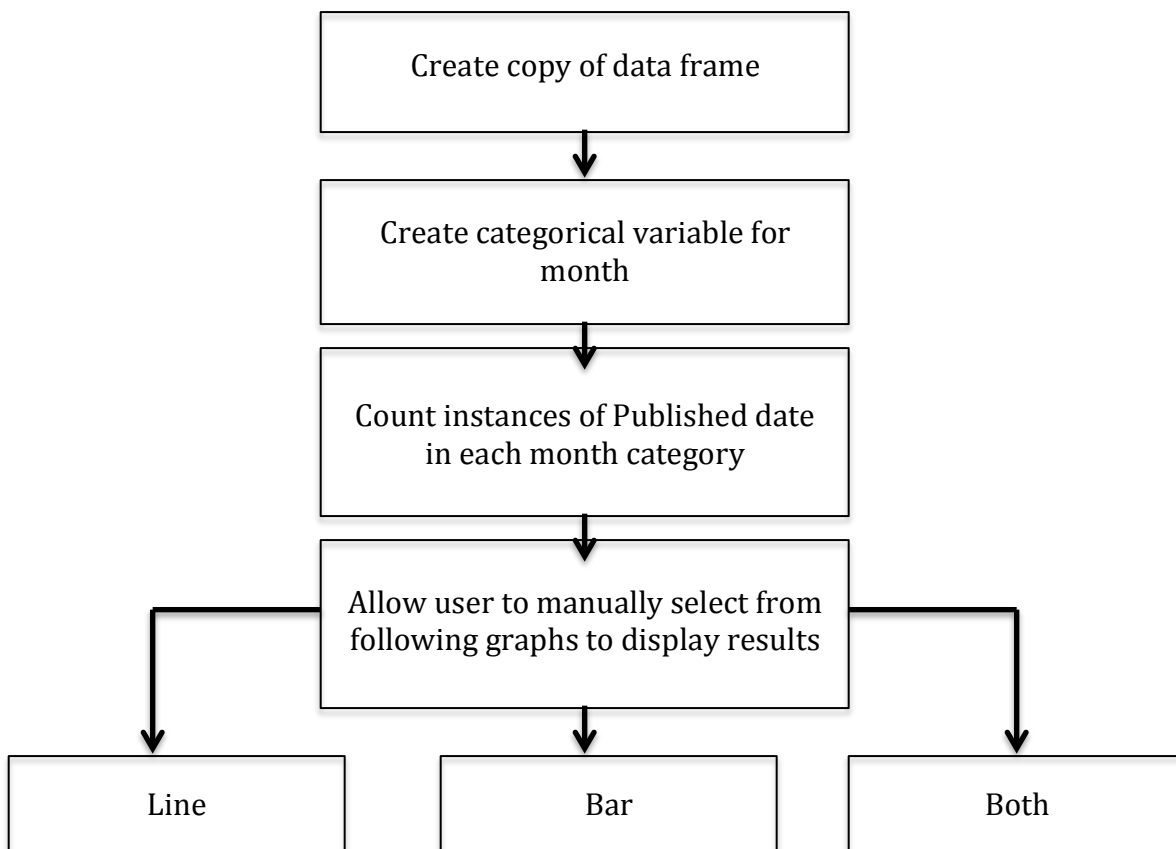


SQL Module



Visualization Module

Draw Graph



Summary Statistics

