**Project Report**

**Project name:** Effect of Global temperature on Sea ice

Our ETL project will give an analysis on effect of global temperature on sea ice. Project goal is to compare maximum and minimum global temperature against the Northern and southern hemisphere of global sea ice.

**Extract:**

We extracted global temperature data from NOAA and sea ice data from National Ice and Snow Data center. Total eight datasets are extracted for global temperature in which two datasets are used for minimum and maximum temperatures and two datasets are extracted for Northern hemisphere sea ice and Southern hemisphere sea ice. Sea ice data is the area of ice covered the sea in square kilometers (sq.km) and maximum and minimum temperatures are in degrees Celsius. Python library pandas is used to read csv files.

**Transform:**

**Sea ice data transform:**

Extracted csv formatted files for both northern and southern hemispheres are converted into Dataframes using Pandas library without a header. Rename the header columns with required column names and set the index as ‘year’ column. Formatted the Date column into date\_time using lambda function. Join the northern hemisphere and southern hemisphere dataframes using ‘merge’. Retrieve the required columns from dataframe into another dataframe. Group the data for the years and extent of ice covered the sea in square kilometers.

**Global temperature data transform:**

Csv files are converted into Dataframes using Pandas library. Renamed the columns as per required and selected the year as index. Un-pivoted the data so that the columns 1 per month are now represented as rows. Converted the month value to string. Added new ‘date’ column using a lambda function. Un-pivoted the data so that the columns 1 per month are now represented as rows and saved each transformed data as a csv file. Collated both the datasets by Year into one dataset for tmax and tmin. Dropped missing rows and transformed tmax and tmin values into degree Celsius.

**Loading:**

SQL Database is used to load the transformed data. Created a database ‘global\_temp\_sea-ice\_db’ and created two tables sea\_ice and global\_temp for sea ice data and global temperature data. Connection is established to MySQL database using SQL Alchemy. Using Pandas ‘.to\_sql’ function data in dataframes are loaded into SQL Database. To view the relationship between global temperature and sea ice join two tables in SQL database.