Impact of Drought on Maple Syrup Production

NOAA - NIDIS Capstone Project

MSBA TEAM B4: Chris Chang, Jacinto Lemarroy, Mengxin Li, Shiyu Ye, Ying Li

Introduction

Maple and the maple industry are synonymous with the New England region's sugar houses and mountainsides with colorful leaves in the fall. Understanding how climate changes and other factors may impact this resource is essential to continue the management of the maple industry into the future.

This project aims to figure out how maple syrup production in gallons is affected by drought and precipitation factors. Our primary focus will be on the Northeast region, specifically seven states, including Massachusetts, Maine, Vermont, New Hampshire, New York, Connecticut, and Pennsylvania. The New England region is well known for its maple syrup production, and hence we will delve deeper into that area. We will analyze variables such as the annual crop production of maple syrup and other parameters related to climate change. Furthermore, this report will further explain the impact of drought on maple tree production.

Methodology

- **Industry Background**
- **Data Description and Source**
- **Data Cleaning and Feature Engineering**
- **Exploratory Data Analysis**
- **Regression Analysis**
- **Conclusion and Recommendations**

Data Collection Analysis

Data Description and Source:

- Collected data from official websites such as the National Oceanic and Atmospheric Administration (NOAA), United States Department of Agriculture (USDA), and National Drought Mitigation Center
- Reached out to organizations such as the New England Forestry Foundation

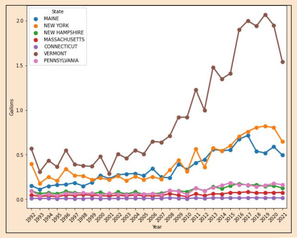
Dataset Summary:

- States: Massachusetts, Maine, Vermont, New Hampshire, New York, Connecticut, and Pennsylvania
- Data timescope: 1992 to 2021
- Target variable: Gallons

Data Cleaning and Feature Engineering

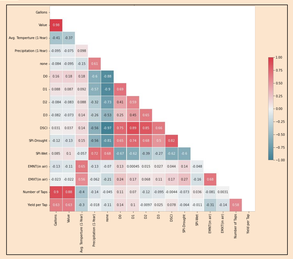
- Missing values in None, D0 - D4, and DSCI columns
- There seemed to be outliers when we plotted boxplots for our variables, but they are real values, so we did not drop them
- Combined existing data with another dataset and used machine learning (OLS and Random Forest) to fill in columns (none, D0 - D4).
- Calculated DSCI column with the following formula: DSCI: 1(D0) + 2(D1) + 3(D2) + 4(D3) + 5(D4)
- Added in SPI (Standard Precipitation Index) columns

Exploratory Data Analysis:



- Vermont, NY, Maine are the three states that produced the most maple syrup in gallons
- Upward trend throughout the years but recent downward tendency

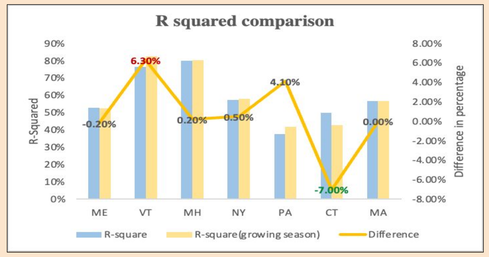
- Gallons, value and number of taps are highly correlated and have a positive relationship.
- Moderate negative correlation between average temperature and both gallons and value
- Drought-related variables are highly correlated



Regression Analysis

We ran the following regressions for our analysis:

- Regression per state
- Regressions on all states
- Lag-effect Regression on all states



- Maple trees growing seasons: precipitation of half year and temperature of all year
- Better using this because R-squared increased overall

Target: Gallons
Predictors: Avg. Temp, Precipitation, D0, D1-D4, DSCI, SPI-Wet, SPI-Drought.

State	R-square	Adjusted R-Squared
ME	52.60%	19.20%
VT	82.70%	70.50%
MH	80.30%	66.40%
NY	58.00%	28.40%
PA	41.70%	0.60%
CT	42.90%	2.60%
MA	56.60%	24.40%
ALL	47.30%	43.80%

- Adjusted R-square is lower.
- Some variables explained less of the model. Potentially deleted variables D1-D3

	Lag effect column	R_squared
One_Year	None	0.459
	Temperature	0.370
	Precipitation	0.467
	Both	0.388
Six_month	None	0.317
	Temperature	0.301
	Precipitation	0.303
	Both	0.272

- None - No lag effect
- Both - Both of temperature and precipitation with lag effect
- Incorporating previous year's precipitation explains more variance in maple syrup production.
- However, this is not the case for average temperature.

Conclusion

- Average temperature has more correlation to gallons compared to precipitation to gallons.
- Specifically, the combination of average temperature (1-year) and precipitation (growing season) correlates most with production in gallons.
- Adjusted R-Squared almost always decreases when we run the regressions with all independent variables.
- When removing variables that are not significant, adjusted R2 increased but R2 decreased since we took out more variables.
- Lag effect indicate that incorporating previous year's precipitation explains more variance in maple syrup production.

Limitations and Challenges

- Lack of samples to run other more sophisticated ML models such as XGboost or Random Forest.
- Difficult to find granular data (monthly data) for maple syrup production.
- Certain variables found online did not have all the data we need (eg. soil moisture data only in county level, DSCI values only available for certain years in certain states)

Potential Improvements

- Add more features (Soil, Geological variables etc.) to improve model.
- Find granular data (monthly, weekly, daily) to perform more accurate analysis.
- Design dashboard (Tableau, PowerBI) for data visualization.
- Create database to store and update data since our data is continuous.
- Expand drought impact analysis on other crops - eg. apples, corn, soybeans.

Acknowledgements

We want to give our gratitude and appreciation to Sylvia Reeves and Dr. Oliver Bandte, who supported us every step of the project. Our project could not have been possible without their guidance and knowledge of the maple syrup industry and the analytics field.