Executive Summary

The objective of this project was to assess the current performance of U.S. businesses using the most recent Business Trends and Outlook Survey (BTOS) data, identify trends and patterns across industries and states, and develop a predictive model to forecast future business performance.

Stakeholders

- **Economists** – To gauge current economic conditions in the US.
- **Investors** – To assess the performance of various sectors and make informed investment decisions.
- **Lawmakers** -To create policies that respond to current economic conditions.
- **Labor Force** – To track employment trends and identify industries prioritizing automation over physical labor.

Feature Engineering Methodology

The survey responses were recorded as a value of the percentage of respondents who were surveyed, with all the options adding up to 100%. However , because the answers were categorical variables, to create a predictive model,I had to recode them as numeric variables .

- For questions with responses (Increased, No Change, Decreased)/ (Yes, No, do not know), values were encoded as (+1, 0, −1).

- For Question 3 (Excellent, Above Average, Average, Below Average, Poor), values were encoded as (+2, +1, 0, −1, −2).

To shrink the multiple responses into a single weighted score , I multiplied the numeric codes by their corresponding percentage values , summed and divided them by 100. This resulted in a single weighted score for each question I wanted to include.

Key Findings

Predicting overall business performance

Since my data was divided into different time periods, my goal was to create a time series forecasting model that could use the data from the previous years 2023-2024 , to predict 2025 businesses performance score.I used a Time-based data split to prevent data leakage, ensuring models predict truly unseen future data. I created two models:

- **Base Model** : Uses the previous period's performance score for prediction
- **Comprehensive Model**: Uses all the previous period's features for predictions and use lasso regression to fine tune the comprehensive model

To conclude, even after tuning the comprehensive model via lasso regression, the comprehensive model did not significantly outperform my base model. This may suggest that a business's recent performance score is the strongest predictor for its performance score in the next two weeks . On the other hand, some information might have been lost when I compressed my multiple choices into a single computed weight score. In the future, I would be more interested in working with BTOS survey responses from individual businesses before they are rounded off as percentages.

Clustering states into different economic profiles using kmeans

The kmeans algorithm was used to possibly cluster the states into different economic profiles. The results were three clusters of sizes 13,12, 28 that you could summarise as best performing , high performing and low performing states .

Best-Performing States(13)

Alaska (AK), District of Columbia (DC), Delaware (DE), Hawaii (HI), Maine (ME), Mississippi (MS), North Dakota (ND), New Hampshire (NH), Rhode Island (RI), South Dakota (SD), Vermont (VT), West Virginia (WV), Wyoming (WY)

Characteristics

- Cost of Inputs Inflation: No Net Cost Inflation- Best of all clusters.
- Change in Selling Prices: Stable Prices - Lowest of all clusters.
- Change in Revenue: Least Negative - Highest of all clusters
- Change in Demand : Least Negative Change in Demand) - Highest demand.
- Change in No of Employees: Nearly Neutral) - Most stable labor market.
- Change in No of Hours Worked : Most Stable Hours- Least negative.
- Current AI Adoption - More in comparison to other clusters
- Future AI Adoption- More in comparison to other clusters

Mid-performing States(12)

Alabama (AL), Arkansas (AR), Idaho (ID), Kansas (KS), Kentucky (KY), Louisiana (LA), Montana (MT), Nebraska (NE), New Mexico (NM), Oklahoma (OK), Puerto Rico (PR), Multi-state)XX).

Worst-Performing States(28)

Arizona (AZ), California (CA), Colorado (CO), Connecticut (CT), Florida (FL), Georgia (GA), Iowa (IA), Illinois (IL), Indiana (IN), Massachusetts (MA), Maryland (MD), Michigan (MI), Minnesota (MN), Missouri (MO), North Carolina (NC), New Jersey (NJ), Nevada (NV), New York (NY), Ohio (OH), Oregon (OR), Pennsylvania (PA), South Carolina (SC), Tennessee (TN), Texas (TX), Utah (UT), Virginia (VA), Washington (WA), Wisconsin (WI).

Characteristics

- Cost of Inputs Inflation: Highest Net Cost Inflation- B
- Change in Selling Prices: Highest price hikes

- Change in Revenue: Most Negative Change in Revenue
- Change in Demand : Least Negative Change in Demand) - Highest demand.
- Change in No of Employees: Least stable labor market.
- Change in No of Hours Worked : Least Stable Hours- Most negative.
- Current AI Adoption - Worst in comparison to other clusters
- Future AI Adoption- Worst in comparison to other clusters

To conclude, the kmeans algorithm was successfully able to cluster the states into different economic profiles.However, to successfully create a strong predictive model, it is essential to work with individual business responses.