



Problem Statement: The Business of State(s)

For Datathon 2026, we will focus on the business of state – specifically, the states of the United States.

California has annual tax revenue of ~250 billion dollars – that’s comparable to Toyota’s annual sales. Michigan has annual tax revenue of ~40 billion dollars – which is not too far below NVIDIA. States of United States are mammoth financial entities with substantial revenue streams – mainly from various taxes and fees that states collect.

Taxes and fees collected by states are a direct result of economic activity conducted in the state and economic policies pursued by the state. While economic trends come and go, one thing has remained certain throughout it all: taxes and fees. Beyond generating revenue for government programs, tax rates and policies can determine consumer spending, investments, and overall economic performance.

You are asked to deep dive in 20 years of data provided and analyze the business of state(s).

State Tax Revenue

Consider these headlines:

- [Gambling generated half a billion dollars in tax revenue for state education fund in 2024](#)
- [Expanded Sports Betting Legalization Would Generate Billions in Tax Revenue](#)
- [Marijuana and sports betting driving US state revenues](#)
- [Which states make the most from sports betting? What about lotteries?](#)

State tax revenues are undergoing a transformation. The past year has brought an eventful wave of economic and policy changes on a national scale. Zoom out further to the past decade, where milestones like the rise of ecommerce, the gig economy, and the AI boom have reshaped the US’s economic landscape. But what’s the big picture at the state level? While some states have consistently relied on specific sectors to bolster their economy, others are shifting in tandem with the changing national economic landscape. Consider the diverging fortunes: a growing wave of states turning to

marijuana legalization, contrasted with Delaware's sudden scramble to retain corporate dominance after [Elon Musk's high-profile exit](#).

Elsewhere, take Ohio's attempt at property tax bans and the pushback from schools losing out on funding—or consider Hawaii's newly blocked tourist tax 'Green Fee' intended to finance sustainability efforts. For the 2026 Datathon, you are asked to explore and analyze 20 years of economic indicators and tax data, challenging you to connect the dots between policy and performance to tell the unseen story of the nation's financial health.

For Michigan Ross Datathon 2026, you are provided with data on all 50 states for the past 20 years. The data include tax revenue and economic indicators. You are asked to explore and analyze this data to answer relevant business questions.

Your Task

Your goal is to analyze the provided datasets (described in detail below), potentially in combination with supplementary public information, to ask relevant business or policy questions and answer them. Please note that we are not providing specific question prompts by design. You are asked to explore what questions can be asked and answered using the provided data. Here we are taking a broad view of “business” and who the potential audience for your analysis might be.

You are asked to pose your own question, explore your own solution approach, and provide answers/solutions using the available datasets in the available time. What is important is the insightfulness and depth of your conclusions and analysis. You need not be comprehensive; quality data analysis will be rewarded over the breadth of the question posed. You however must look at your task as broader than just data analysis – **you should look for a business/policy question and consider the business context**. Note that thoughtfully prepared data visualizations are an integral part of your data exploration and analysis.

Creativity in formulating your own well-thought-out question or analysis approach generally has a positive effect on judges' assessment of your submission; however, it should not be at the expense of analytical depth, precision, and rigor, which are far more important. Please note that the provided datasets contain a lot of information – and you need not use all or even most of it for your chosen business question. You may decide to focus on a specific narrow slice/dice of the dataset that truly speaks to you and go deeper in your chosen narrow focus.

Your analysis should be well communicated to a non-technical audience, potentially with the help of appropriate visualizations.

Given that Ross students typically do not interact extensively with macroeconomic/government data, here are a few starting lines of enquiry that can be asked using this dataset:

- Does Florida's efforts to phase out property taxes make sense? How will Florida provide essential services to its residents without the property tax revenue. [See details of the legislative effort here](#).
- Many states rely on State Income Taxes while many other states do not have any Income Tax. Is there any difference in economic performance between these two different kind of state taxation regimes?

- Have new tax sources like legalization of Gambling and Marijuana made a significant difference to state tax revenues?
- Which states are in a relatively better fiscal position and which states are suffering? What policy advice would you have for states?

These are just starting thoughts. You are encouraged to think deeply about the goals, objectives and the purpose of state government – and decide on a suitable question to ask and answer. Note that the datasets provides are extensive and it is not necessary for you to use all the provided data. You have data for all 50 states but you may decide to focus your analysis efforts on one or a few states. Creativity and imagination in asking a good, relevant and impactful question is the most important contributor to a great submission.

You are asked to keep your focus on ONLY state governments and not the federal government.

Datasets

Datasets are provided in Excel/CSV format. CSV files can be easily opened with all data analysis tools including MS Excel. Please note that the datasets are **somewhat cleaned** but you may still need to do some cleaning and transformation before you can proceed to data analysis.

For source of the dataset, see the section about state tax revenue here: <https://www.data-is-plural.com/archive/2024-06-12-edition/>

You have two sets of data:

1. US States Tax Revenue Data, Quarterly, for last 20 years
2. State wise Economic Indicator Data, Quarterly, for last 20 years

Caution: you must first spend enough time to understand the dataset. Datathon Canvas discussion forums are a great place to ask and answer questions about the data. **Please note that you need not use every bit of data provided – use the part of the data that speaks to you, that sparks curiosity in you, that lends insights to you.**

Additional Datasets

You are welcome to look for additional datasets/information to supplement your analysis. All additional data used should be public and their source must be documented in the submission. Even though you can use additional dataset to supplement the provided dataset, **you must keep the provided dataset central to your analysis.** While evaluation of your submission will take a holistic approach and evaluate your entire submission, **we will consider whether you have done justice to the provided dataset or not.**

Data Dictionary

While the economic indicators have self-explanatory column headings, the tax types in the state revenue data needs explanation. See explanation of the data dictionary provided below.

Tax Description	Code
Total Taxes	
Property taxes	T01
Sales and gross receipts taxes	
General sales and gross receipts	T09
Selective sales and gross receipts taxes	
Motor fuels	T13
Alcoholic beverages	T10
Public utilities	T15
Insurance premiums	T12
Tobacco products	T16
Sports betting (including pari-mutuels)	T18
Amusements	T11
Other selective sales and gross receipts	T19
License taxes	
Alcoholic beverages	T20
Public utilities	T27
Motor vehicles	T24
Motor vehicle operators	T25
Corporations in general	T22
Hunting and fishing	T23
Amusements	T21
Occupation and businesses	T28
Other license taxes	T29
Income taxes	
Individual income	T40
Corporation net income	T41
Other taxes	
Death and gift	T50
Severance	T53
Documentary and stock transfer	T51
Other taxes, NEC	T99

NEC: Not Elsewhere Classified

Further details of the data dictionary is provided in a separate file.

Submission: Content

You are expected to submit two separate deliverables: a recorded presentation of max 8 min duration and a presentation file. Your submissions must be in easily accessible formats (e.g PPT/PDF for presentation file, mpeg for recorded video). You can use any tool like Zoom for creating your recorded presentation video.

Additional information that may not fit the presentation time may be placed in an appendix of the presentation file. You will be asked to make your submissions to a Canvas assignment. Note that deadlines are strictly enforced and any significant delay in submission will result in your submission being rejected.

Judges will be evaluating your work without your team there to explain it; therefore, your submission must “**speak for itself**”. It need not be polished to the level of a final product, but do ensure that your main findings are clear and that any visualizations are functionally labeled and interpreted.

Note that since we don’t provide a prompt, you are free to choose your audience. You may look at the data from the point of view of emissions, firms, customers, regulatory bodies, or any other stakeholders. However, you should clearly document your audience and the value you seek to provide to them.

You are asked to build and submit a presentation file - this is the file you should have used for your recorded presentation – and you will use to present your solution live to the expert panel of judges should you be chosen as one of the 6 finalist teams.

Tips & Recommendations

We recommend that you follow a low-hanging-fruit first approach. While it is tempting to go for big and bold ideas, remember that you have limited time and even essential tasks like data cleaning and joining tables are likely to take significant time. We strongly recommend that you take care of low, effort-high return tasks first before going for more in-depth analytics. In particular, effective data visualization is an efficient, high return, low(er) effort target to finish first. It is important that you do not get stuck in a complex technical workflow and lose track of not just the time but also your overall analysis objective. If something is taking too long, then it is okay to shelve it for the moment and move on to more easily executable tasks. Prioritize simple, easy, doable tasks first.

We recommend that your team not try to learn new tools if possible; instead, leverage your existing skills to extract as much insight from the data as you can. You will find that even basic tools like Excel and Tableau, if used well, can be quite effective in a time bound competition like this where speed is essential. Further, a good business narrative and interpretation from a simple analysis is likely more valuable than complex technical analysis that does not lend itself to (or does not leave time for) a compelling business insight.

Note that your submission will be evaluated for its overall communication, structure, and narrative not just on your technical work. **If your submission does not adequately communicate the story coming out from your technical work then all your work is essentially wasted.** Your submission should present a compelling story and that story should be clear even with a cursory look at your submission.

We STRONGLY encourage you to start building up your final submission AT LEAST two to three hours before the submission deadline. In the past, many teams have spent a lot of time conducting great analyses, only to realize that they left almost no time for actually writing up and presenting their results. This cannot be stressed enough – quality data analysis that is incomplete or poorly presented will NOT win the competition.

Potential Analysis Approaches

You have a lot of data and no fixed decision problem to solve. In such situation it is recommended that you keep in mind the following:

1. Before you do any “analysis”, you would want to explore the data – just get to know what you have, how are the variables distributed, whether significant cleaning is needed. A through

exploration of the data will not only show you any data cleaning need, it will also point towards insights that can form the basis for later solutions.

2. Once you have a sense of the data, you would want to get the two important questions sorted:
 - a. Who is your audience? Who are you doing your analysis for? For the dataset provided, your audience could be State Legislatures, Policy Analysts, Macroeconomic Consulting, Journalists etc. Your analysis can't solve everything for everyone – so step 1 is to choose who is going to be your audience.
 - b. What problem are you solving for your audience? Your solution must be rooted in a need. You must provide some value for your audience. Identify the problem you are addressing and through your analysis, show how you will provide value for your audience.
3. Now that you have an audience and a problem, you would want to tell an effective story that will connect with and persuade your audience to make the needed decisions. All your analysis is worthless if that does not move the needle for your audience – and people are persuaded by stories and visuals – not by numbers and regressions for example. Tell an effective story.

Ask for Help

Do not hesitate to ask for help. You will have access to a Canvas discussion board where you can ask questions and the Datathon organizers will respond as appropriate. Representatives from sponsor companies will be offering Office Hours where you can get their advice. Datathon organizing team will be available throughout the day in case you have any technical or logistical issues – just let us know and we will do our best to assist.

If you wish to ask a question privately (i.e. not in the public Canvas Discussion Thread) then please send an email to Sanjeev (sankum@umich.edu).

That's it. We hope that you will have a wonderful experience in the Michigan Ross Datathon 2026 and that it will prove a fun learning experience for you.