Data excise tax

- Conclusion
  - A monthly excise tax on commercial data collectors using the approach and rates proposed in S.2012 of 2023 could raise $700-800 million annually when fully and effectively implemented - perhaps even more. Higher or lower tax rates could be imposed than those in S.2012.
  - Even if the tax does not face significant legal risks, there are many revenue estimating risks because (1) analytic-firms' estimates of unique visitors vary widely, and (2) data collectors might have legitimate ways to reduce the number of reported unique visitors.
  - I'd suggest an estimate for financial planning purposes much lower than the top of the range, perhaps $350-400 million, to allow for the uncertainties.
- Option summary
  - "Commercial data collector" (CDC):
    - **For-profit** entity that collects, maintains, uses, processes, sells or shares consumer data in support of business activities.
    - Collects consumer data "other than consumer contact information" on **more than 1 million individual NY consumers in a month**.
      - Other-than-consumer-contact-information -- data on device IDs, devoce type, browser, demographics, preferences, transactions, movies watched, pages visited, songs listened to, questions asked, purchases, ...
      - Contact data sold to third parties treated as non-contact data
    - Data collectors with common ownership would be combined -- e.g., Alphabet (Google, YouTube and YouTubeTV, Waymo,...), Meta (Facebook, Instagram, WhatsApp, ...), Amazon (Amazon Web Services, Twitch, Audible, Whole Foods, Zappos), Microsoft (LinkedIn, GitHub, Skype, Xbox)
  - Consumer
    - An *individual* who purchases or uses goods or services from CDC. The single member of a single-member LLC is a consumer.

        - NY consumer has primary residence in NY per NYS income tax residency rules.
        - A NY consumer counts only once per month per data collector, regardless of:
            - Whether their data is collected by multiple platforms of the data collector (e.g., Facebook and Instagram in the case of Meta)
            - How many ways they engage with platforms (phone, computer, gaming device, TV set box, etc.)
        - A NY consumer could be included in tax base of multiple CDCs in a month
    - Graduated tax rates based on # of NY consumers a CDC collects data on in a month:
        - Like an income tax: base amount plus marginal rate x # consumers in range
        - Analysis uses rates of S.2012 of 2023 (but alternative rates could be imposed):
            - < 1 million NY consumers in month: 0 cents per consumer
            - 1-2 million consumers: tax is $0 + 5 cents per NY consumer in range (tax on 2 million is $50k), ranging up to...
            - > 10 million consumers: $2.25 million + 50 cents per NY consumer (tax on 12 million NY consumers in a month is $2.25m + (50 cents x 2 million) = $3.25million -- annual rate of $39 million, or $3.25 per NY consumer per year.)
    - CDC and Tax Department may agree on methodology for determining NY consumers
    - CDC may claim a credit for tax paid to another state that imposes identical excise tax on the same NY consumer.
- Methodology
    - Estimating goals
        - Obtain/estimate data for each large commercial data collector on the number of unique NY individuals about which it collects data in a typical month.
        - Data should include unique individuals only once in a month for companies that have common ownership -- the Alphabet (Google, Youtube) example above.

- The data should count unique individuals only once in a month regardless of how they access the data collector's sites, or how many times.
- The data should exclude purely non-individual consumers -- e.g., the commercial restaurant that uses a Spotify account to play background music.

- Key assumptions
  - **Comscore data on U.S. monthly unique visitors for top 50 platforms is a good starting point** (December 2024 Digital Media US, Home and Work)
  - **80% of national unique visitors are individuals** -- judgmental based on my review of multiple web sites and articles and querying of several chatbots; Comscore seeks to track unique individuals (100%) so 80% could be conservative
  - **100% of national unique monthly visitors have non-contact data collected from them**: have not found any data on this; presumably platforms are almost always collecting non-contact data from individuals
  - **5.8% of national unique individuals are NY residents** -- population share
  - **Visitors to platform 100 are 70% of platform 50, and 51-100 average is average of #50 and #100** We have no data on platforms 51-100. #20 was 70% of #10 and #40 was 68% of #20. Assume #100 is 70% of #50. Calculate #51-100 tax using average of #50 and #51 visitors.
  - **50m monthly unique NY resident visitors to non-online-platform data collectors**: Many private firms that are not internet platforms could become taxpayers because they collect data from individuals -- banks and other financial institutions, mobile and landline phone companies, hospitals and health care networks, private utility companies, private universities, and more.
    - The top 5 financial institutions or financial services firms (JPMorgan Chase, Bank of America, Citibank, Wells Fargo, and Capital One) all serve more than 2 million NY residents and serve about 27 million in total, allowing duplication across firms.
    - The top 5 telecommuncations firms serving NY residents (Verizon, Spectrum, T-Mobile, Optimum) each serves more than 1 million NY residents and in aggregate they appear to serve more than 25 million customers, allowing duplication across firms.

    - The top 5 health care networks serving NY residents (Northwell, NYU Langone, Mount Sinai, New York-Presbyterian, and Montefiore) each serve more than 1 million NY residents (smallest, Mount Sinai, is about 1.4 million) with a total of about 8.4 million, allowing duplication across firms.
    - The top 3 privately owned utilities (ConEd, National Grid, and NYSEG) all have more than 1 million NY customers and likely more than 8 million in total after combining electric and gas and allowing for some within-firm overlap, and allowing duplication across firms.
    - These numbers developed with help of Deepseek chatbot.
    - Presumably most of the above collect some kind of information from individuals in most months.
    - In combination, these companies likely have likely have more than 50 million NY customers (allowing duplication). In addition, other important industries and firms that could be taxpayers are not counted in the numbers above.
    - This adds about $29 million to the estimate before various risks
  - **50% data, implementation, and legislative risk** -- Given huge uncertainties I think it's prudent to budget only half or so of the estimate. Data risk is that we don't know the number of unique monthly individuals residing in NY from whom data are collected, deduplicated across devices and platforms with common ownership. Different analytics firms have significantly different estimates. Implementation risk is the risk that data collectors find large loopholes that allow them to change how they gather, use, or report information in a way that drastically reduces, for tax purposes, the number of users they report while still allowing them to report large numbers of users for advertising and similar purposes. Legislative risk is the risk that the legislation will be watered down substantially in negotiations - perhaps exempting industries or ownership types.
- **Results**: These assumptions lead to an estimate of about $785 million before allowing for avoidance, data, implementation, or legislative risks. I suggest only including about half of the estimate for financial planning purposes - perhaps $350-400 million. See this spreadsheet, cells N2 and P1 and related notes for calculations

- **Note**: To see if it's possible to get better data I sent the following request to Comscore. We'll see if they disclose useful information or if they have a reasonable cost proposal (I think both unlikely): "I am interested in obtaining, for a recent representative month, the number of unique visitors to platforms such as those found in your Comscore December 2024 Digital Media US, Home and Work. I'm looking for data on individual visitors (not business visitors); limited to platforms that collect at least some data beyond contact information; that are deduplicated across platforms with common owners (e.g., an individual who visits Google and YouTube would count only once, for Alphabet); deduplicated across number of times accessed (10 visits to YouTube from an individual in a month count as one visitor); deduplicated across modes of access (phone, tablet, computer, settop box, etc.); where the individuals are residents of New York state. I know your data can't do this exactly, but I'm wondering whether you could provide me with something close, and tell me what it would cost, and also help me understand the ways in which it might fall short of what I am requesting."
- **Illustration of estimation risks:**
  - Analytics firms use 2 basic methods of estimating unique visitors
    - Direct tracking of site usage via cookies and certain non-cookie methods (e.g., Google Analytics)
    - Nielsen-like panels of users (e.g., Comscore)
  - Both have weaknesses - direct tracking can do a lot of double counting using cookies. Panels of users often extrapolate estimates from relatively small samples of users.
  - Both have run into problems over the years (I can provide links):
    - Facebook notoriously overestimated video views for years, according to WSJ stories, was sued for it in 2018
    - Podcast downloads have been overstated -- e.g., "A user who listened to a show like The New York Times' The Daily a few times, subscribed, but stopped listening would continue to count as a download indefinitely."
    - Comscore relied on "users installing a data collection tool on their browser to enter Nielsen-style user panels and that tool isn't compatible with any Apple system or the latest version of Windows"

- Comscore (perhaps self-servingly) argued that cookie-based tracking (like Google Analytics) likely overestimates visitors by 20-30%; users can delete cookies and be counted again. Comscore claims about 30% of internet users delete cookies monthly averaging 4 deletions per month and can overstate unique visitors by a factor of 2.5
    - Users clearing cookies (40% do this monthly)
    - Multiple device usage (50% of Americans use 4+ internet devices)
    - Browser changes
    - Incognito mode usage
- One academic paper concluded that SimilarWeb consistently counted fewer unique visitors than Google Analytics, averaging 38.7% lower
- Comcast says there is a trend toward not obtaining IDsS: "more than half of all mobile impressions (54%) and over a third of desktop impressions (36%) no longer contain a user identifier, including alternative identifiers. This lack of addressability underscores the urgency for marketers to adopt privacy-centric, ID-free solutions to maintain scale in their digital ad campaigns." Not sure yet what this means:
    - One reaction - almost all marketers now use cookies
    - Not sure yet what this means. Even if marketers don't get ID info, presumably the tax would still be due. By whom?
- I attempted to spot check several companies. This is hard to do because there is little free data on measures that align with the tax base. Some analytics sourcesprovide unique visitors, some provide page views or other traffic measures, some analytics sources do not deduplicate across devices, some only provide worldwide measures, and more. My rough sense is that alternative data sources