



## Frequently Asked Questions: UQ Election Ad Data Dashboard

---

### **Q. Where does the data come from?**

Ads are collected from the Facebook Ad Library API (<https://www.facebook.com/ads/library>) which contains sponsored posts declared by the advertiser as political. We collect the settings for *each ad* as they currently are at the time of retrieval from the ad library, many of which can be modified over the time period that an ad runs on the platform, such as the estimated spend on the ad. This often results in differences between what our dashboard outlines is spent on an ad and what is listed in the Facebook Ad Library as this uses a combination of invoices as well as estimates.

### **Q. How should our data be interpreted?**

Due to the nature of the data, our dashboard cannot provide exact numbers on spend. It is a useful tool to analyze patterns in spend over time, party, terms, regions, and demographics. But when reporting our figures, it is important to note the limitations and caveats with the data.

Collecting the data at the granularity of the ad level allows us to provide much richer analytics at the expense of some differences in the exact spend. If you require exact spend figures for a page, please consult the Facebook Ad Library.

### **Q. How do we process the Data?**

To analyse the data, we apply a number of pre-processing steps. This involves manual annotations of ads performed by crowdsourcing on Amazon MTurk as well as supervised machine learning models trained with such manual annotations. The machine learning models together with pre-trained natural language processing are used to extract information such as aspect-based sentiment of the posts, key terms and topics discussed within each ad. Party, Electorate, and Candidate labels are extracted from the ad's page name and funding entity and are matched with Australian Electoral Commission (AEC) candidate, electorate, and party data as well as via manual labeling and verification. We are currently tracking ad spend for close to 4000 pages.

### **Q. How often does the data update?**

We gather fresh data every 6 hours. The dashboard shows data at a day-level granularity.

### **Q. What is on the dashboard?**

The dashboard contains an overview of the data we are collecting and generating. We are unable to show data for every electorate or every candidate. We will aim to disseminate as much additional data and observations as we can during the campaign via public posts and on social media.



### **Q. What does the underlying data pipeline look like?**

The system runs on Amazon AWS. Every 6 hours we collect active ad campaigns from the Facebook Ad Library and enrich it with in-house AI tools to extract additional information. We then store it on S3. From here the data is served via Elasticsearch, DynamoDB, and Athena to a number of clients for data analytics including Tableau, RStudio, and PowerBI. The dashboard takes this data and uses plotly to generate the visualizations.

### **Q. How is the ad spend estimated?**

Data provided by Facebook only includes spend ranges (e.g., \$0-\$99) rather than actual spend amounts. This requires us to estimate the actual spend of an ad campaign. Options include taking the upper-bound (e.g., \$99) in all cases, or the average value (e.g., \$50) for each ad. This may lead to over or under estimation of the actual spend. For example, this is likely for small ad campaigns which are short in duration and have few impressions or where there are many ads running.

Based on observations in regards to the accuracy of these spends, we define the total spend on an ad as the most recent upper bound spend for an ad. For example, if the ad says the spend is between 0 and 99, we use 99 as the spend. Such an approach will overestimate short-lived ads for which the actual spend is closer to the lower bound of the range.

To spread the total spend on an ad over the entire period it is running, the spend per day of an ad is calculated by taking the total spend and dividing it by the total days the ad has been / was active. Therefore, each day an ad is active, it will contribute total spend/total days active.

### **Q. How is the party spend figure calculated?**

This figure shows all ad spend by any page associated with a party each day. This includes candidate pages, state and national party pages, as well as local branch or regional funding entity spend. For example, for Labor, this includes the spend from the Opposition Leader, Anthony Albanese's page, senators not up for re-election, the state Labor Party pages, the Australian Labor Party national page, local branches spending money on ads, as well as the pages of individual candidates contesting this election. It does not include associated entities for any party such as trade unions. We have this data but do not show it on the dashboard.

It is important to recognise that as this is a 'real-time' dashboard and the fact the Meta advertising system is dynamic (so that over the lifecycle of an ad the estimated spend can be adjusted by the funding entity), it is impossible to know how much is going to be spent when an ad initially becomes active. Hence, we suggest caution is taken with these numbers and that they are seen as indicative of a broader strategy.

### **Q. How have you calculated the spend in each electorate?**

We use the ad spend from the candidate pages as a proxy for electorate spend. For the purposes of consistency, we have removed the leaders of the parties from the spend calculation for their electorates as these spends are not going into these electorates. Despite these limitations, we think this is still useful for showing the spend of candidates in different electorates around the country.



Note: Do not report these as actual spend in the electorate!

**Q. How have you calculated the Senate spend?**

While this is a half-Senate election, the pages of the Senators not up for re-election are still spending on ads. Thus, we have included the spend from the pages of all the current Senators, as well as all the candidates for the Senate in this election.

**Q. How have you calculated the trending data?**

These tables primarily show the parties and terms that correspond to the greatest increase in spend for the current week.

The "Spend This Week" columns are calculated from Monday until the current day of the week. The "Spend Past 4 Weeks" columns are calculated from the Monday of 4 weeks prior until the current day. The "Weekly Spend Change" is calculated by determining the percentage change between the spend this week, and the average spend over the past four weeks.

Within the "Trending Terms" table, the "Party Spend This Week" corresponds to how much the listed party has spent on that term. The "Total Term Spend" is the sum of spending across all parties during the timeframe.

This table removes any instances of parties talking about themselves, or their respective leaders, to put the focus on other terms and "negative advertising" throughout the campaign.

**Q. If I have questions about the data and project, who can I contact?**

Dr. Gianluca Demartini - [g.demartini@uq.edu.au](mailto:g.demartini@uq.edu.au) (Data and Computing Infrastructure)

Dr. Glenn Kefford - [g.kefford@uq.edu.au](mailto:g.kefford@uq.edu.au) (Political Commentary and Media)