## **Exhibition Demand Prediction Model**

Worksheet Model [here](https://royalacademy-my.sharepoint.com/:x:/r/personal/xiaoxiao_zeng_royalacademy_org_uk/Documents/analysis_env/project/Predicting%20Model/2024%20Exhibition%20Demand%20Projections%20-%20%20Updated%20Model%2020241219.xlsx?d=w358bdb4761dc4fe589e25d279077a176&csf=1&web=1&e=HPMGmY)

Github resipotary and Google Colab Code reference here

### **1. Key variables**

This model predicts **how many people will attend an exhibition** based on:

* **Google search volume**
* **Duration** (number of days the exhibition runs)
* **Press coverage**
* **Gallery**
* **Era** (pre-COVID, COVID, or post-COVID)
  + The reason to include this is that we have to use historical data to build the model. And not including this factor will bias the model.

### **2. Search Volume**

**How is search volume calculated?**Search volume refers to the google search volume using tools like Google trends or third-party tools like [search volume](https://searchvolume.io/). For different types of exhibitions, there are different ways to do this, depending on how the exhibition is positioned:

**1. Artist’s name recognition directly**

* Some exhibitions are driven by well-known artists, so the search volume is based purely on the artist's name.
* Example:
  + Marina Abramović exhibition → Search volume: 6,186

**2. Combining artist's name and exhibition theme**

* If an artist is less widely known, but the exhibition has a strong **theme**, we use a mix of both.
* Example:
  + Angelica Kauffman exhibition → Combined search volume: 1,737
  + Calculation:
    - Angelica Kauffman (highly relevant, weighted 90%) → Search volume: 425
    - Women artists (less relevant, weighted 10%) → Search volume: 13,552
    - Final search volume = (425 × 0.9) + (13,552 × 0.1) = 1,737

**3. Using only themes**

* Some exhibitions are positioned around broader art themes rather than individual artists.
* Example:
  + "Eye of the Storm" exhibition → Combined search volume: 8,067
  + Calculation:
    - Ukraine art (highly relevant, weighted 80%) → Search volume: 4,033
    - Modernism (less relevant, weighted 20%) → Search volume: 24,200
    - Final search volume = (4,033 × 0.8) + (24,200 × 0.2) = 8,067

The way search volume is calculated depends on how we **position the exhibition**—whether we emphasize the artist’s fame, the exhibition’s theme, or a combination of both.

### **3. Press coverage**

Press coverage is used to determine if an exhibition has received great attention in the press. To determine which exhibition is and which not, we need to find a threshold. When the press coverage hits a certain number, we can label it as a high press coverage exhibition.

**How do we find the threshold?**Since press coverage varies between the main gallery and smaller galleries, we need to decide **what counts as "high" press coverage** for different gallery types.

1. We calculate percentiles of press coverage separately for main and small galleries.
2. We test different cutoff points (e.g., top 50%, top 60%, etc.) to find the best definition of "high" coverage.
3. For each cutoff, we check how well the model predicts attendance.
4. We choose the cutoff that makes the predictions most accurate.

Think of it like baking cookies:

* If you bake them for too little time, they’re undercooked.
* If you bake them for too long, they burn.
* The goal is to find the perfect baking time where they turn out just right.
* In our case, we test different thresholds for "high press coverage" and choose the one that makes our attendance predictions most reliable.

**Caveat:**

* Fewer exhibitions in small galleries reach the level of "high press coverage" compared to main galleries.
* This means that while we define high press coverage for GJW and Sackler, only a few exhibitions actually meet this threshold in practice and result will be biased.

### **4.How different factors affect attendance**

* **Search Volume:**
  + A 10% increase in search volume leads to 3.3% more visitors.
  + Example: If an exhibition originally attracts 100,000 visitors, a 10% increase in search volume raises attendance to 103,300.
* **Number of Days:**
  + A 10% increase in exhibition duration leads to 12.8% more visitors.
  + Example: If an exhibition runs for 50 days and gets 100,000 visitors, extending it to 55 days increases attendance to 112,800.
* **Extending an exhibition doesn’t always mean more people come per day**.
  + If an exhibition is extended from 30 days to 60 days, total attendance will increase, but daily average attendance might drop.
  + Suppose an exhibition originally had 3,000 visitors per day over 30 days (total: 90,000 visitors).
  + If extended to 60 days, the total might rise to 115,200 visitors, but the daily average could drop to 1,920 visitors per day.
* **Press coverage has a strong impact**
  + High press coverage → +29.9% attendance
  + Example: If an exhibition with low press coverage gets 40,000 visitors, strong media attention could raise attendance to 51,600 visitors.
  + Caveat: Since fewer small gallery exhibitions reach high press coverage, the impact of high press coverage in small galleries is based on a limited number of cases.
* **Main Gallery has higher baseline attendance than GJW and Sackler**
* **The average attendance differs between Pre-Covid, Covid and Post Covid.**

**5. How to implement the model**

The current model assumes a baseline attendance of 19,135 visitors (Post Covid) when:

* **An exhibition lasting ~90 days**
* **A monthly search volume of ~1,800 searches**
* **A small gallery (GJW or Sackler) with low press coverage**
* **GJW**

Using this model, we can know how the predicted attendance will change accordingly when different factors change.

1. **Search volume**

* Baseline search volume: 1,800 searches → 19,135 visitors
  + If search volume increases to 3,600 (+100%) → Attendance ≈ 25,500 visitors (+33%)
  + If search volume drops to 900 (-50%) → Attendance ≈ 15,900 visitors (-17%)

1. **Number of days**

* Baseline duration: 90 days → 19,135 visitors
  + If duration increases to 135 days (+50%) → Attendance ≈ 31,414visitors (+64%)
  + If duration decreases to 45 days (-50%) → Attendance ≈ 6,841 visitors (-64%)

1. **If the exhibition was in a Main Gallery:**

* Attendance increases by ~445% (or 545× the baseline).
* New Attendance: 19,135 × (1 + 445%) = ~12,320 visitors

1. **If it had high press coverage:**

* Attendance increases by ~29.9%.
* New Attendance: 19,135 × (1 + 54%) = ~24,859 visitors