# Quantifying Movie Magic with Google Search

mazefeng01@baidu.com

Baidu, Inc

2014 年 5 月 9 日

# 目录

# Introduction

# Introduction

- This paper discuss how search query patterns and paid clicks can help us to predict box office performance.

# Introduction

- This paper discuss how search query patterns and paid clicks can help us to predict box office performance.

- Searches in the movie category on Google are up 56% from 2011 to 2012.

# Introduction

- This paper discuss how search query patterns and paid clicks can help us to predict box office performance.

- Searches in the movie category on Google are up 56% from 2011 to 2012.

- Search provides a unique opportunity for film marketers to extend engagement with potential moviegoers.

# Introduction

- This paper discuss how search query patterns and paid clicks can help us to predict box office performance.

- Searches in the movie category on Google are up 56% from 2011 to 2012.

- Search provides a unique opportunity for film marketers to extend engagement with potential moviegoers.

- Query volume and paid click volume along with other movie-related variables, can predict:

# Introduction

- This paper discuss how search query patterns and paid clicks can help us to predict box office performance.

- Searches in the movie category on Google are up 56% from 2011 to 2012.

- Search provides a unique opportunity for film marketers to extend engagement with potential moviegoers.

- Query volume and paid click volume along with other movie-related variables, can predict:
    - A film's opening weekend performance with 92% accuracy.

# Introduction

- This paper discuss how search query patterns and paid clicks can help us to predict box office performance.

- Searches in the movie category on Google are up 56% from 2011 to 2012.

- Search provides a unique opportunity for film marketers to extend engagement with potential moviegoers.

- Query volume and paid click volume along with other movie-related variables, can predict:
  - A film's opening weekend performance with 92% accuracy.
  - Subsequent weekend performance with 90% accuracy.

# Introduction

# Introduction

- Trailer(预告片)-related search trends 4 weeks out from a movie release provide strong predictive power for opening weekend box office revenue.

# Introduction

- Trailer(预告片)-related search trends 4 weeks out from a movie release provide strong predictive power for opening weekend box office revenue.

  - Trailer search volume on Google coupled with both the franchise status of the movie(经销权) and seasonality can predict opening weekend box office revenue with 94% accuracy.

# Introduction

- Trailer(预告片)-related search trends 4 weeks out from a movie release provide strong predictive power for opening weekend box office revenue.
    - Trailer search volume on Google coupled with both the franchise status of the movie(经销权) and seasonality can predict opening weekend box office revenue with 94% accuracy.

- 48% of moviegoers decide what film to watch the day they purchase their ticket.

# The Link Between Google Search and Box Office

If search is a reflection of interest and intent, one would expect that the more movie-related search activity in a given weekend, the bigger the box office.

Figure 1 | Comparison of 2012 Box Office Index and Film-Related Search Index
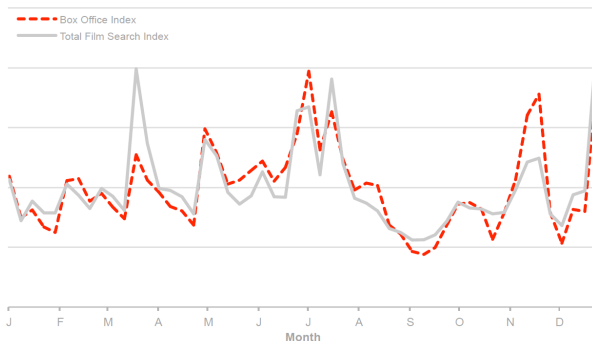
# The Link Between Google Search and Box Office

If we split the overall search index into film-specific title keywords and generic movie-related keywords:

# The Link Between Google Search and Box Office

If we split the overall search index into film-specific title keywords and generic movie-related keywords:

- Film-specific title keywords include "[movie title]", "[movie title] trailer", "[movie title] clips", "[movie title] cast", "[movie title] tickets", "[movie title] plot", "[movie title] reviews", and other common variants related to these categories.

# The Link Between Google Search and Box Office

If we split the overall search index into film-specific title keywords and generic movie-related keywords:

- Film-specific title keywords include "[movie title]", "[movie title] trailer", "[movie title] clips", "[movie title] cast", "[movie title] tickets", "[movie title] plot", "[movie title] reviews", and other common variants related to these categories.

- Generic movie-related keywords include general movie terms (e.g. "new movies", "movie showtimes"), theater chain terms (e.g., "regal showtimes", "carmike theaters"), and online movie ticket services (e.g., "fandango", "movietickets").

# The Link Between Google Search and Box Office

Figure 2 | Comparison of 2012 Box Office Index, Movie-Title Search Index, and Non-Title Film-Related Search Index
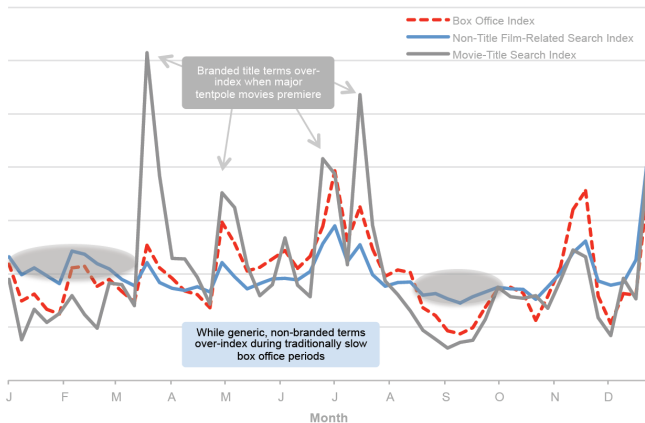
# Predicting Weekend Box Office One Day Before: Paid Clicks Predict Tix

# Predicting Weekend Box Office One Day Before: Paid Clicks Predict Tix

- 99 films released in 2012 shows that a film's Google search volume during the 7 days prior to its opening is a strong indicator.

# Predicting Weekend Box Office One Day Before: Paid Clicks Predict Tix

- 99 films released in 2012 shows that a film's Google search volume during the 7 days prior to its opening is a strong indicator.

- A simple linear regression model using film-related search query volume as a predictor of weekend box office performance yields an $R^2$=70%.

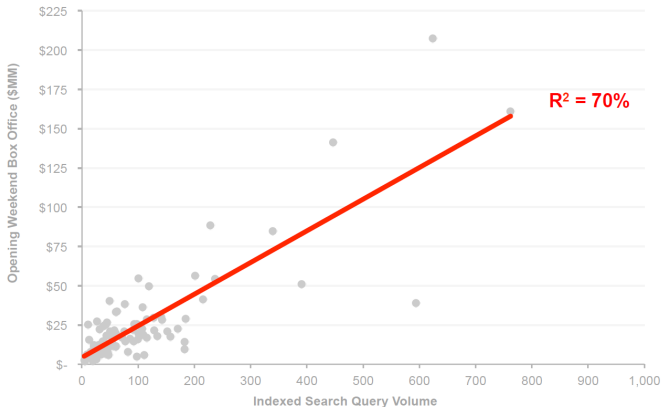# Predicting Weekend Box Office One Day Before: Paid Clicks Predict Tix

- 99 films released in 2012 shows that a film's Google search volume during the 7 days prior to its opening is a strong indicator.

- A simple linear regression model using film-related search query volume as a predictor of weekend box office performance yields an $R^2$=70%.

- In other words, 70% of the variation in box office performance can be explained with search query volume.

# Predicting Weekend Box Office One Day Before: Paid Clicks Predict Tix

Figure 3 | Scatterplot of Indexed Search Query Volume and Actual Box Office Opening Weekend



R² = 70%

# Predicting Weekend Box Office One Day Before: Paid Clicks Predict Tix

# Predicting Weekend Box Office One Day Before: Paid Clicks Predict Tix

- **Search ad click volume** is a key element of the prediction model as it signifies deeper engagement with a particular film.

# Predicting Weekend Box Office One Day Before: Paid Clicks Predict Tix

- Search ad click volume is a key element of the prediction model as it signifies deeper engagement with a particular film.
- After examing over 30 different film variables, non-statistically significant variables were dropped, resulting in a model that includes the following variables:

# Predicting Weekend Box Office One Day Before: Paid Clicks Predict Tix

- **Search ad click volume** is a key element of the prediction model as it signifies deeper engagement with a particular film.
- After examing over 30 different film variables, non-statistically significant variables were dropped, resulting in a model that includes the following variables:
    1. Search query volume (seven-day period prior to release date)

# Predicting Weekend Box Office One Day Before: Paid Clicks Predict Tix

- **Search ad click volume** is a key element of the prediction model as it signifies deeper engagement with a particular film.
- After examing over 30 different film variables, non-statistically significant variables were dropped, resulting in a model that includes the following variables:
  1. Search query volume (seven-day period prior to release date)
  2. Search ad click volume (seven-day period prior to release date)

# Predicting Weekend Box Office One Day Before: Paid Clicks Predict Tix

- **Search ad click volume** is a key element of the prediction model as it signifies deeper engagement with a particular film.
- After examing over 30 different film variables, non-statistically significant variables were dropped, resulting in a model that includes the following variables:
  1. Search query volume (seven-day period prior to release date)
  2. Search ad click volume (seven-day period prior to release date)
  3. Theater count

# Predicting Weekend Box Office One Day Before: Paid Clicks Predict Tix

- Search ad click volume is a key element of the prediction model as it signifies deeper engagement with a particular film.
- After examing over 30 different film variables, non-statistically significant variables were dropped, resulting in a model that includes the following variables:
  1. Search query volume (seven-day period prior to release date)
  2. Search ad click volume (seven-day period prior to release date)
  3. Theater count
  4. Franchise(经销权) status

# Predicting Weekend Box Office One Day Before: Paid Clicks Predict Tix

Figure 4 | Opening Weekend Prediction Model (Search Ad Click Volume as X-axis)

# Predicting Weekend Box Office One Day Before: Paid Clicks Predict Tix

# Predicting Weekend Box Office One Day Before: Paid Clicks Predict Tix

- This model has an $R^2$ of <span style="color:red">92%</span>, meaning box office sales can be predicted by the combination of the aforementioned variables with 92% accuracy.

# Predicting Weekend Box Office One Day Before: Paid Clicks Predict Tix

- This model has an $R^2$ of 92%, meaning box office sales can be predicted by the combination of the aforementioned variables with 92% accuracy.
- In the 7 day window prior to a film's release date:

# Predicting Weekend Box Office One Day Before: Paid Clicks Predict Tix

- This model has an $R^2$ of 92%, meaning box office sales can be predicted by the combination of the aforementioned variables with 92% accuracy.
- In the 7 day window prior to a film's release date:
  1. If one film has 250,000 more search queries than a similar film, the film with more queries is likely to perform up to $4.3M better.

# Predicting Weekend Box Office One Day Before: Paid Clicks Predict Tix

- This model has an $R^2$ of 92%, meaning box office sales can be predicted by the combination of the aforementioned variables with 92% accuracy.
- In the 7 day window prior to a film's release date:
  1. If one film has 250,000 more search queries than a similar film, the film with more queries is likely to perform up to $4.3M better.
  2. If a film has 20,000 more paid clicks than a similar film, it is expected to bring in up to $7.5M more.

# Predicting Weekend Box Office One Day Before: Paid Clicks Predict Tix

# Predicting Weekend Box Office One Day Before: Paid Clicks Predict Tix

- Beyond Opening Weekend: the number of paid clicks a film garners along with other film-related metrics, serve as strong indicators of how a film will fare in a post-opening weekend.

# Predicting Weekend Box Office One Day Before: Paid Clicks Predict Tix

- Beyond Opening Weekend: the number of paid clicks a film garners along with other film-related metrics, serve as strong indicators of how a film will fare in a post-opening weekend.
- For holdover films (i.e. films in their second week or beyond), the significant variables in our regression model include:

# Predicting Weekend Box Office One Day Before: Paid Clicks Predict Tix

- **Beyond Opening Weekend**: the number of paid clicks a film garners along with other film-related metrics, serve as strong indicators of how a film will fare in a post-opening weekend.
- For holdover films (i.e. films in their second week or beyond), the significant variables in our regression model include:
  1. Search ad clicks (Monday-Thursday prior to weekend)

# Predicting Weekend Box Office One Day Before: Paid Clicks Predict Tix

- **Beyond Opening Weekend**: the number of paid clicks a film garners along with other film-related metrics, serve as strong indicators of how a film will fare in a post-opening weekend.
- For holdover films (i.e. films in their second week or beyond), the significant variables in our regression model include:
  1. Search ad clicks (Monday-Thursday prior to weekend)
  2. Theater count

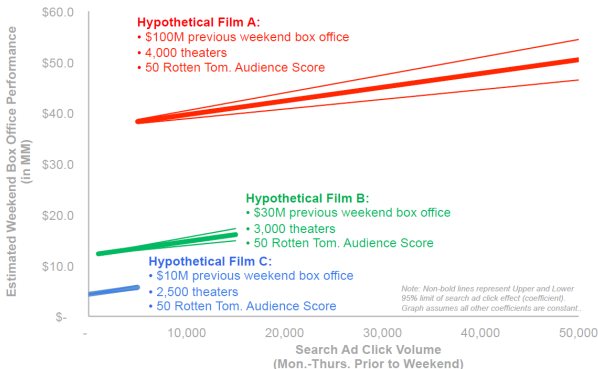# Predicting Weekend Box Office One Day Before: Paid Clicks Predict Tix

- **Beyond Opening Weekend**: the number of paid clicks a film garners along with other film-related metrics, serve as strong indicators of how a film will fare in a post-opening weekend.
- For holdover films (i.e. films in their second week or beyond), the significant variables in our regression model include:
  1. Search ad clicks (Monday-Thursday prior to weekend)
  2. Theater count
  3. Previous weekend performance

# Predicting Weekend Box Office One Day Before: Paid Clicks Predict Tix

- Beyond Opening Weekend: the number of paid clicks a film garners along with other film-related metrics, serve as strong indicators of how a film will fare in a post-opening weekend.
- For holdover films (i.e. films in their second week or beyond), the significant variables in our regression model include:
  1. Search ad clicks (Monday-Thursday prior to weekend)
  2. Theater count
  3. Previous weekend performance
  4. Rotten Tomatoes audience score

# Predicting Weekend Box Office One Day Before: Paid Clicks Predict Tix



Figure 5 | Holdover Weekend Prediction Model (Search Ad Click Volume as X-axis)

# Predicting Weekend Box Office One Day Before: Paid Clicks Predict Tix

# Predicting Weekend Box Office One Day Before: Paid Clicks Predict Tix

- Search query volume for holdover films is not a significant factor in predicting box office performance after opening weekend, while search ad clicks remain a strong indicator.

# Predicting Weekend Box Office One Day Before: Paid Clicks Predict Tix

- Search query volume for holdover films is not a significant factor in predicting box office performance after opening weekend, while search ad clicks remain a strong indicator.
  - Once a film has opened, search ad clicks are a strong sign of intent to purchase a ticket.

# Predicting Weekend Box Office One Day Before: Paid Clicks Predict Tix

- Search query volume for holdover films is not a significant factor in predicting box office performance after opening weekend, while search ad clicks remain a strong indicator.
  - Once a film has opened, search ad clicks are a strong sign of intent to purchase a ticket.
  - The intent associated with a search query is more varied.

# Predicting Weekend Box Office One Day Before: Paid Clicks Predict Tix

- Search query volume for holdover films is not a significant factor in predicting box office performance after opening weekend, while search ad clicks remain a strong indicator.
  - Once a film has opened, search ad clicks are a strong sign of intent to purchase a ticket.
  - The intent associated with a search query is more varied.

- If one film garners 10,000 more paid clicks is likely to perform approximately 1.9−3.5M better.

# Predicting Box Office One-Month Before Release: In Trailer Searches We Trust

# Predicting Box Office One-Month Before Release: In Trailer Searches We Trust

- Leave movie marketers with very much time to react and adjust their marketing campaigns in the pre-release time frame.

# Predicting Box Office One-Month Before Release: In Trailer Searches We Trust

- Leave movie marketers with very much time to react and adjust their marketing campaigns in the pre-release time frame.

- Overall title search volume loses some of its predictive power the further it is from premiere date.

# Predicting Box Office One-Month Before Release: In Trailer Searches We Trust

- Leave movie marketers with very much time to react and adjust their marketing campaigns in the pre-release time frame.

- <span style="color:red">Overall title search volume loses some of its predictive power</span> the further it is from premiere date.

- The key to long-range box office forecasting lies in movie trailer engagement, <span style="color:red">trailer-related search query volume</span> holds strong predictive power.

# Predicting Box Office One-Month Before Release: In Trailer Searches We Trust

# Predicting Box Office One-Month Before Release: In Trailer Searches We Trust

- The number of trailer searches mattered most in terms of predicting box office four weeks prior to a film's release week ($R^2 = 62\%$).

# Predicting Box Office One-Month Before Release: In Trailer Searches We Trust

- The number of trailer searches mattered most in terms of predicting box office four weeks prior to a film's release week ($R^2 = 62\%$).

- Similar to trailer-related Google searches, title-related searches on YouTube have the highest predictive power four weeks from release date ($R^2 = 55\%$).

# Predicting Box Office One-Month Before Release: In Trailer Searches We Trust

- The number of trailer searches mattered most in terms of predicting box office four weeks prior to a film's release week ($R^2 = 62\%$).

- Similar to trailer-related Google searches, title-related searches on YouTube have the highest predictive power four weeks from release date ($R^2 = 55\%$).

- 94% of variation in a film's box office opening can be explained with trailer-related title search volume 4 weeks prior to release, coupled with seasonality and franchise status.

# Predicting Box Office One-Month Before Release: In Trailer Searches We Trust



Figure 6 | Predictive Power of Trailer-Related Search Volume and Overall Search Volume