ECON 0150 | Economic Data Analysis

The economist's data analysis pipeline.

Part 1.7 | Grouping Data

Grouping Data: Starbucks Promotions

Q. Which promotional offers change behavior the most?

- *We have a dataset `starbucks_promotions.csv` with individual events (offers sent, transactions made)*
- *Each row represents one event with an Offer ID, Event type, and Revenue*
- *Raw data shows thousands of individual events*

> *looking at individual data won't easily tell us which offers work best…*

> *we need to **group** the data to see patterns*

Grouping Data: The Process

Transform individual rows into group summaries.

1. ***Filter*** *(if needed) - keep only relevant rows*
2. ***Group*** *- organize rows by categorical variables*
3. ***Aggregate*** *- summarize each group with statistics*

> this transforms many rows into fewer, more meaningful summaries

Common Aggregation Functions

What can we calculate for each group?

Function	Purpose	Example Use
`.sum()`	Total values	Total revenue per offer
`.count()`	Number of rows	Number of events per offer
`.mean()`	Average values	Average transaction amount
`.max()`	Maximum value	Highest single transaction
`.min()`	Minimum value	Lowest transaction amount

> *choose the aggregation that answers your question*

Starbucks Offers

Q. Which promotional offers change behavior the most?

1. How likely is each offer-type to be used?

- *How many times was each offer-type sent?*
- *How many times was each offer-type used?*

2. What's the average revenue per offer-type?

- *What is the total revenue for each offer-type?*
- *How many times was each offer-type sent?*
- *What is the average revenue for each offer-type?*

Starbucks Offers: Grouping and Counting

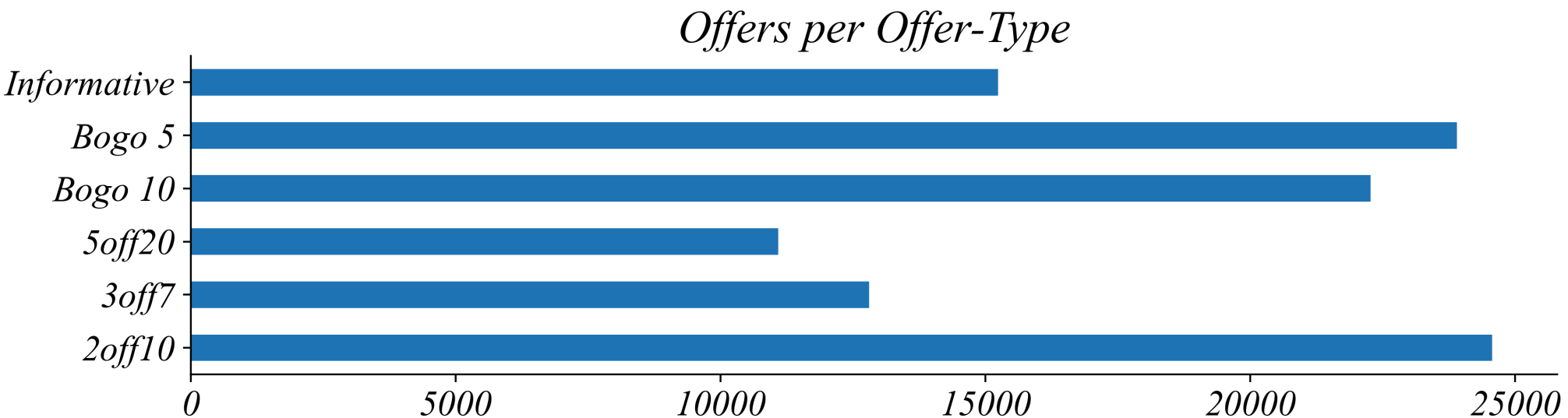
How frequent is each offer-type?

Lets start by counting each offer-type in the ***categorical variable*** `Offer ID`.

Starbucks Offers: Grouping and Counting

How frequent is each offer-type?

Lets start by counting each offer-type in the ***categorical variable*** `Offer ID`.



> *so Bogo 5 and 2off10 get sent most frequently*

> *but it also looks like there are multiple types in `Event'*

Starbucks Offers: Grouping and Counting

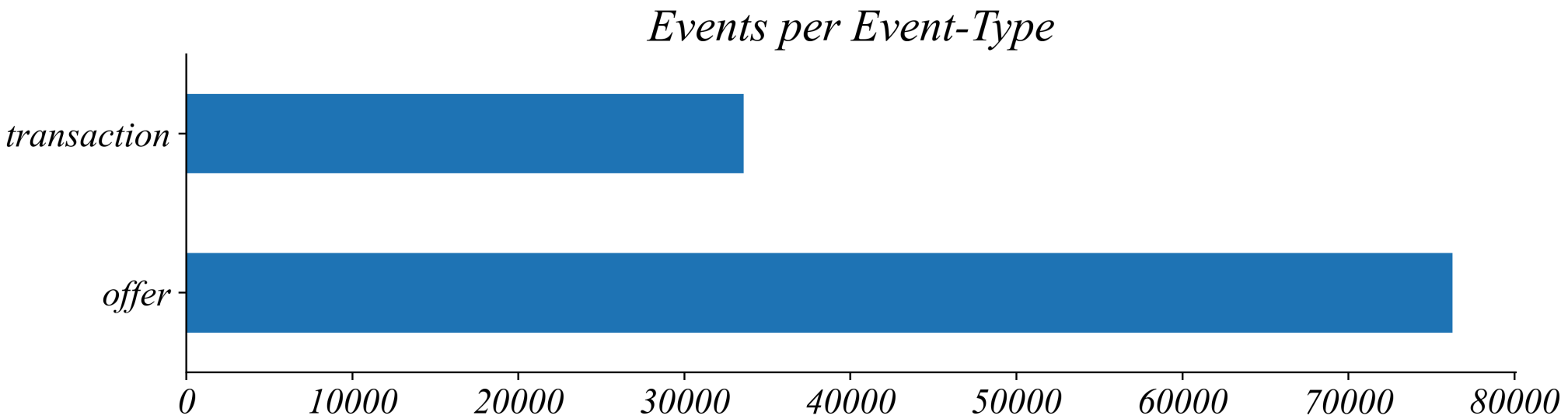
How frequent is each offer-type?

Lets also group on `Event` to count event-types.

Starbucks Offers: Grouping and Counting

How frequent is each offer-type?

Lets also group on `Event` to count event-types.



> it looks like 'offer' may not actually be a transaction... lets investigate

Starbucks Offers: Grouping and Summing

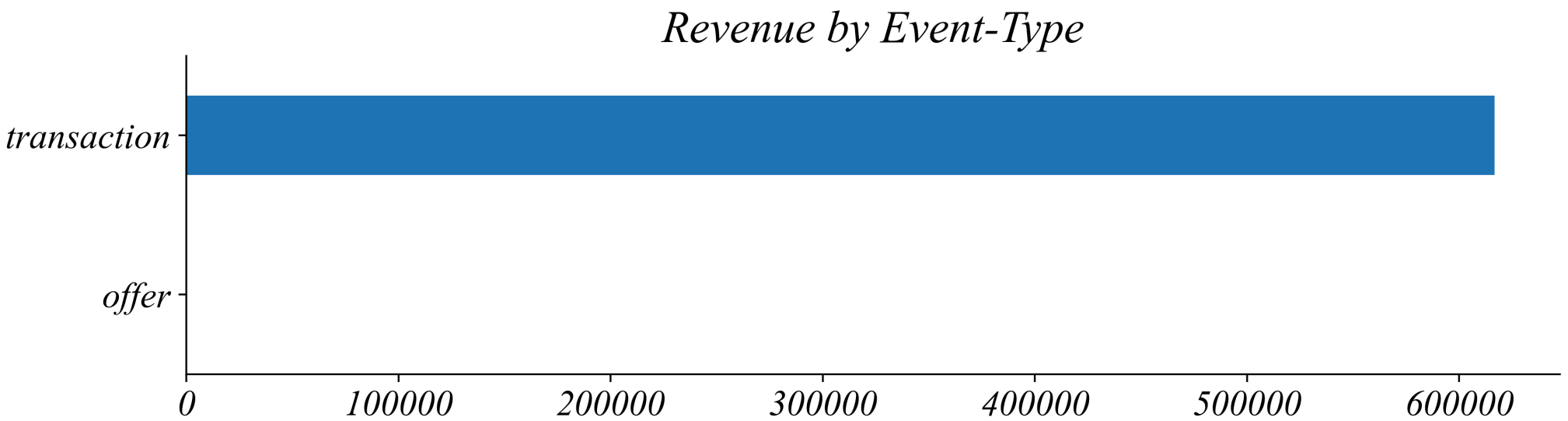
What is the total revenue for each event-type?

Lets group on `Event` and find the total revenue by event-type.

Starbucks Offers: Grouping and Summing

What is the total revenue for each event-type?

Lets group on `Event` and find the total revenue by event-type.



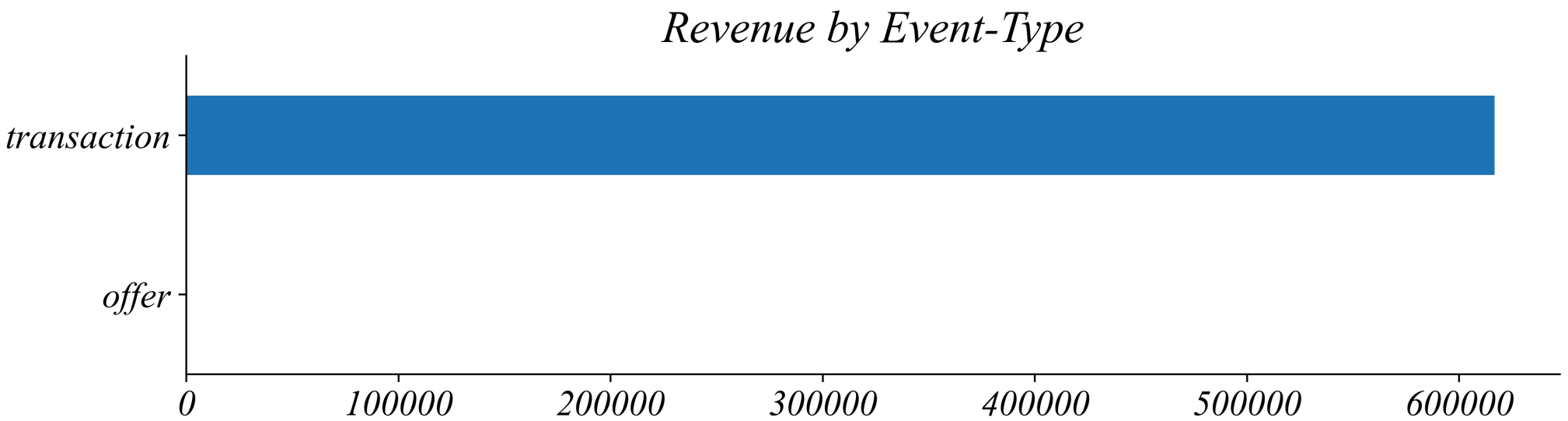
> *just as we suspected, "offer" events brings in 0 revenue*

> *so* `Event` *tells us whether the row is an 'offer' or a 'transaction'*

Starbucks Offers: Grouping and Summing

What is the total revenue for each event-type?

Lets group on `Event` and find the total revenue by event-type.



> lets filter on `Event` to break this down by `Offer ID`

Starbucks Offers: Grouping and Counting

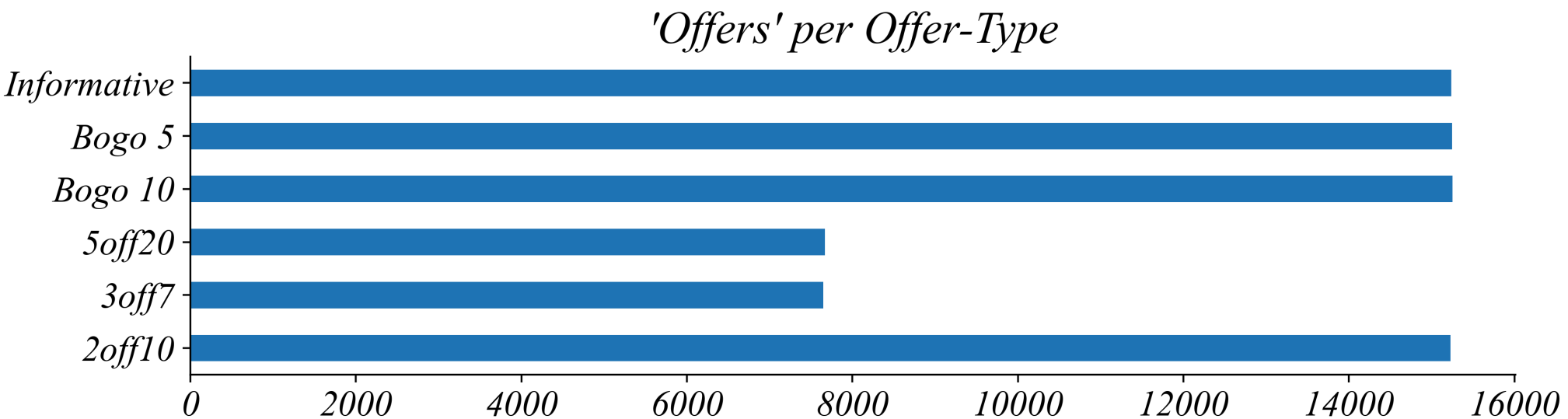
How many times was each offer-type sent?

Lets filter `Event` for just ‘offer’ then group on `Offer ID` and count.

Starbucks Offers: Grouping and Counting

How many times was each offer-type sent?

Lets filter `Event` for just 'offer' then group on `Offer ID` and count.



> *so Bogo 5 and 2off10 were sent most frequently*

> *lets check how often each offer-type was used*

Starbucks Offers: Grouping and Counting

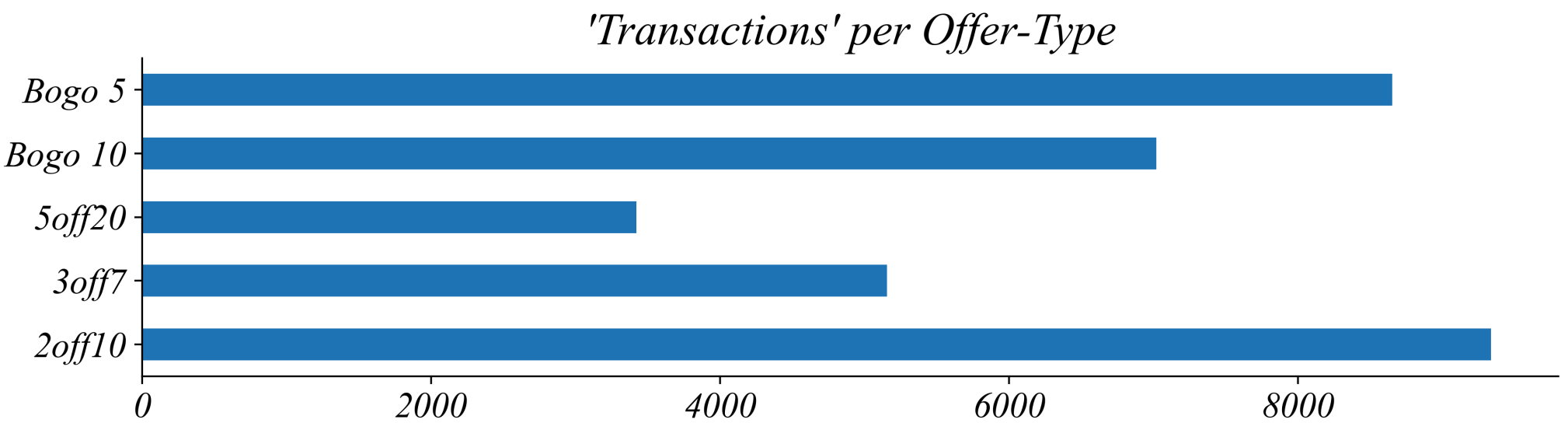
How many times was each offer-type used?

Lets filter `Event` for just ‘transaction’ then group on `Offer ID` and count.

Starbucks Offers: Grouping and Counting

How many times was each offer-type used?

Lets filter `Event` for just ‘transaction’ then group on `Offer ID` and count.

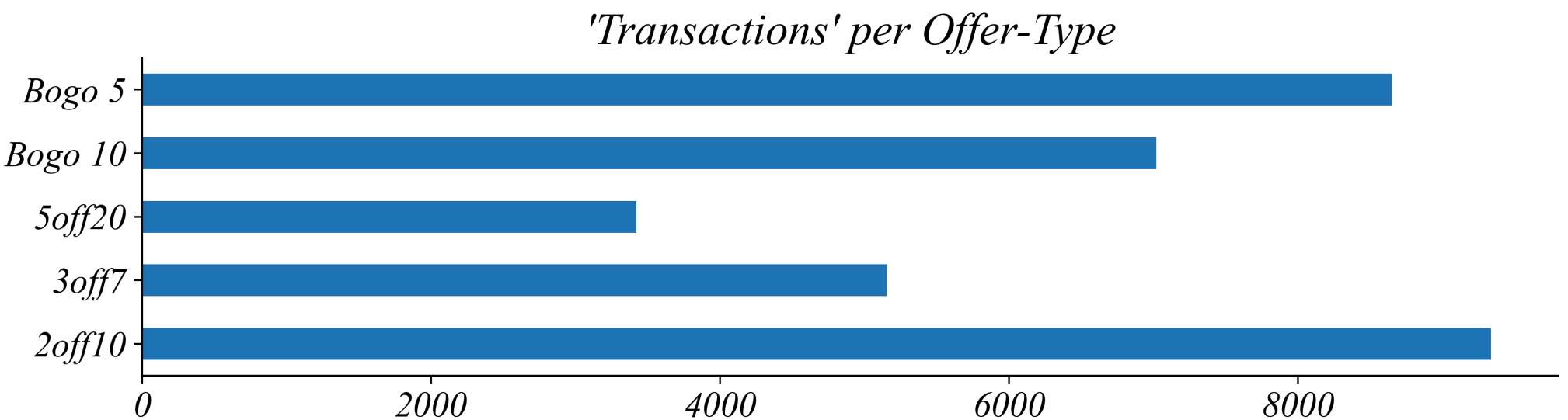


> *so Bogo 5 and 2off10 were used most frequently*

Starbucks Offers: Grouping and Counting

How many times was each offer-type used?

Lets filter `Event` for just ‘transaction’ then group on `Offer ID` and count.



> *does this mean they were the most effective?*

> *no! we want to find how likely an offer will turn into a transaction when sent*

Starbucks Offers: Simple Transformations

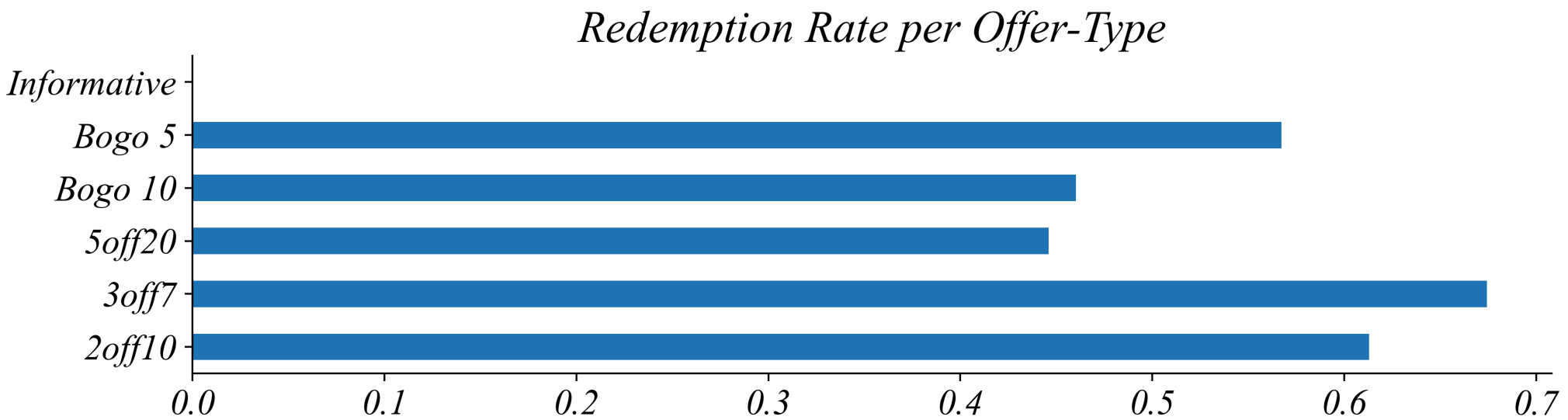
How likely is each offer-type to be used?

Lets divide the number of transactions by the number of offers sent by offer-type.

Starbucks Offers: Simple Transformations

How likely is each offer-type to be used?

Lets divide the number of transactions by the number of offers sent by offer-type.

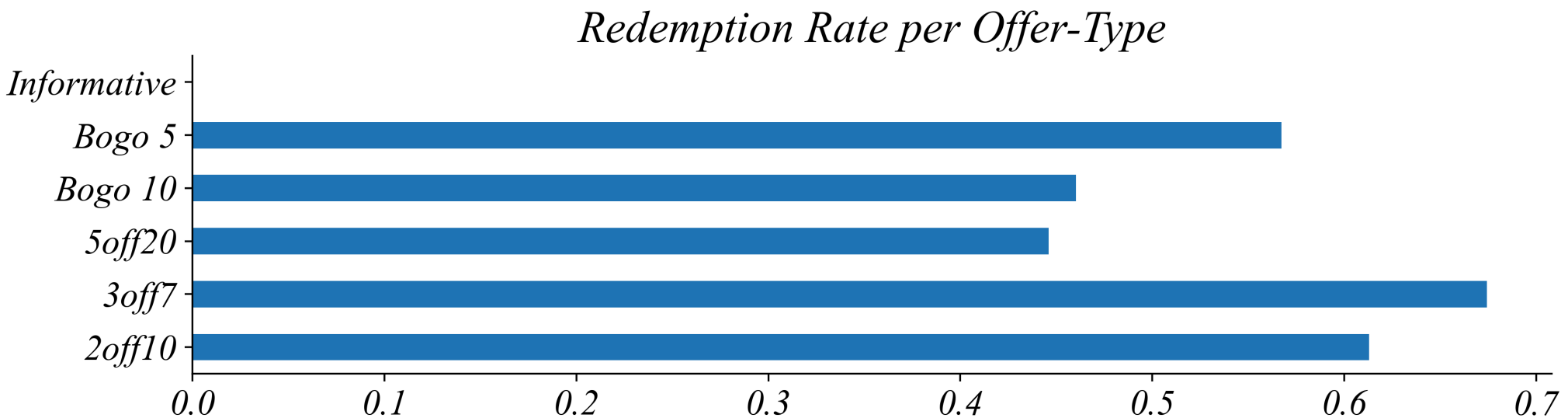


> *Bogo 5, 3off7, and 2off10 have the highest use rate!*

Starbucks Offers: Simple Transformations

How likely is each offer-type to be used?

Lets divide the number of transactions by the number of offers sent by offer-type.



> *this is a pretty reasonable way to measure 'effectiveness'*

> *but how much we can expect to bring in each time an offer is sent?*

Starbucks Offers: Grouping and Summing

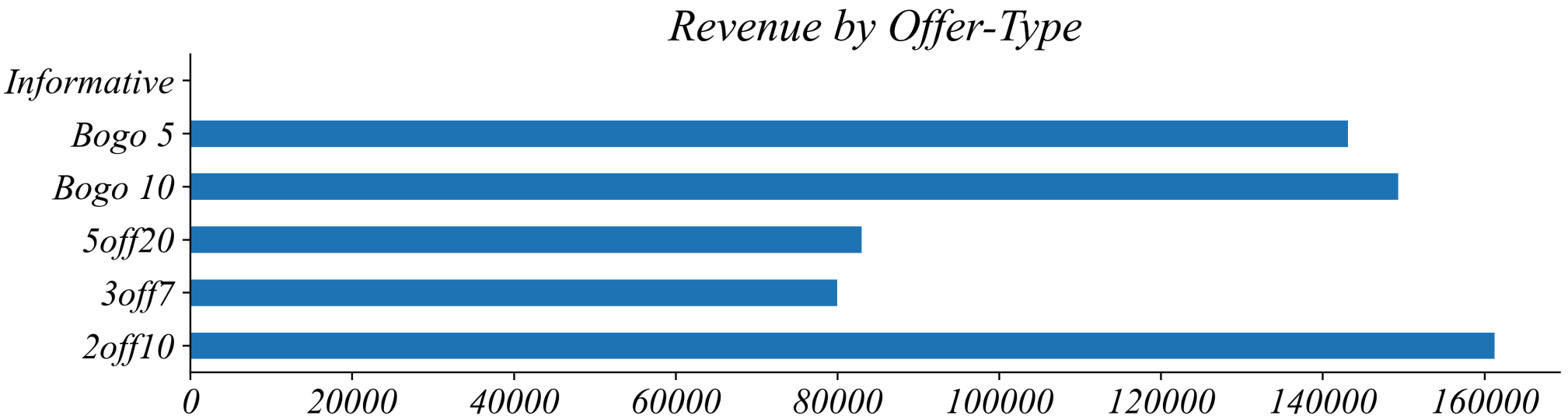
What is the total revenue for each offer-type?

Before finding average revenue by `Offer ID`, lets start by finding total revenue.

Starbucks Offers: Grouping and Summing

What is the total revenue for each offer-type?

Before finding average revenue by `Offer ID`, lets start by finding total revenue.



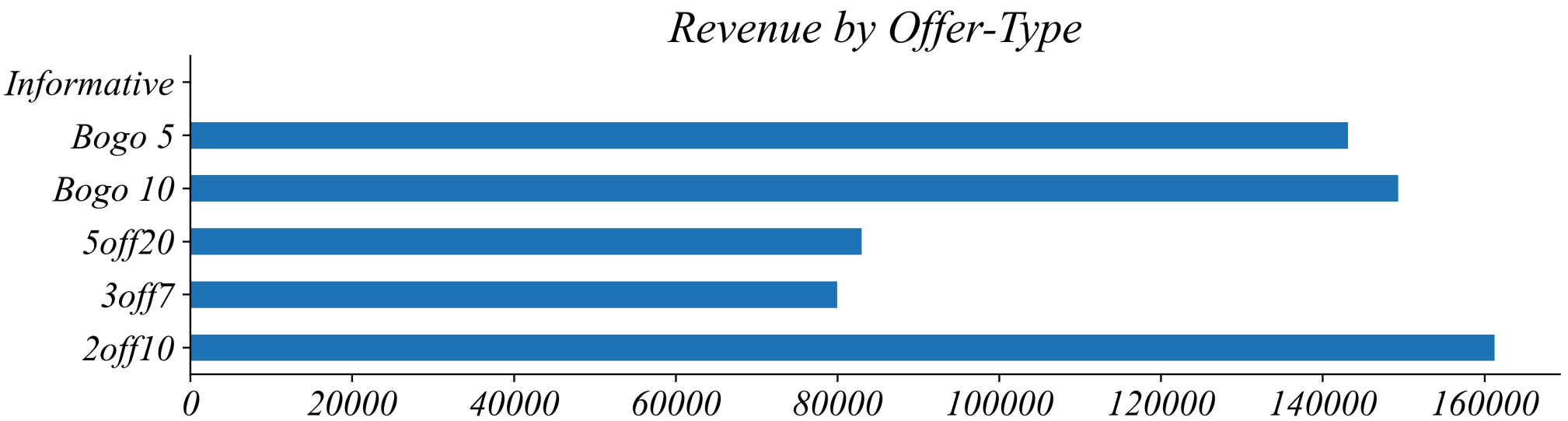
> *'Informative' brings in no revenue (*makes sense*)*

> *Bogo 5, 3off7, and 2off10 bring in the most revenue*

Starbucks Offers: Grouping and Summing

What is the total revenue for each offer-type?

Before finding average revenue by `Offer ID`, lets start by finding total revenue.



> *does this mean 2off20 is the best promotion?*

> *not necessarily… lets find the average revenue by offer-type*

Starbucks Offers: Grouping and Averaging

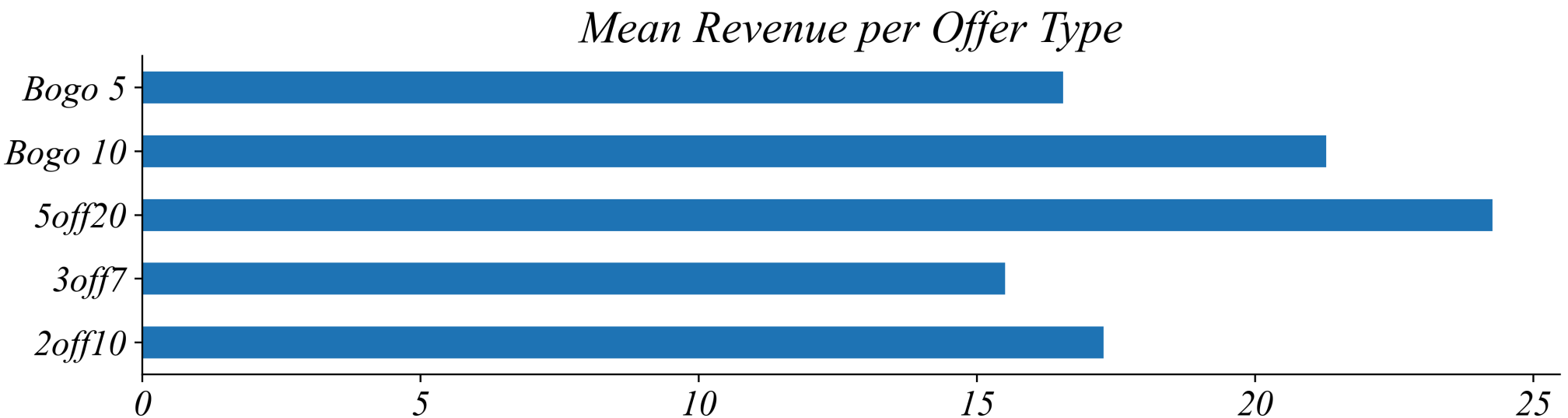
What is the average transaction amount per offer type?

Lets find the average (mean) transaction amount by offer-type.

Starbucks Offers: Grouping and Averaging

What is the average transaction amount per offer type?

Lets find the average (mean) transaction amount by offer-type.



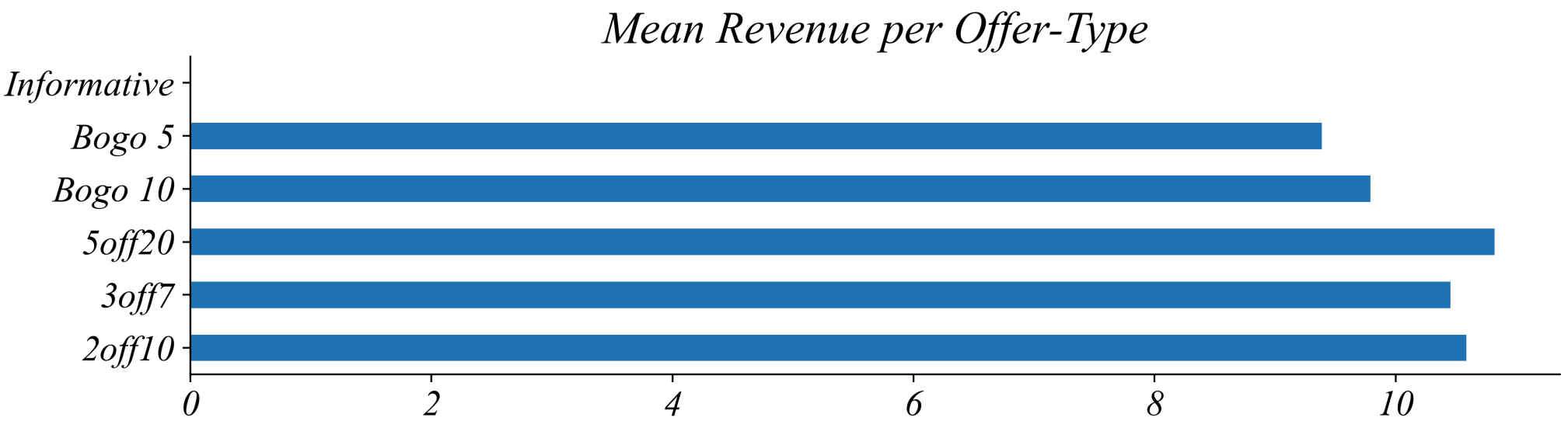
> *this tells us how much people spend per transaction if they use the offer*

> *it does not tell us how much revenue we can expect after sending an offer*

Starbucks Offers: Grouping and Averaging

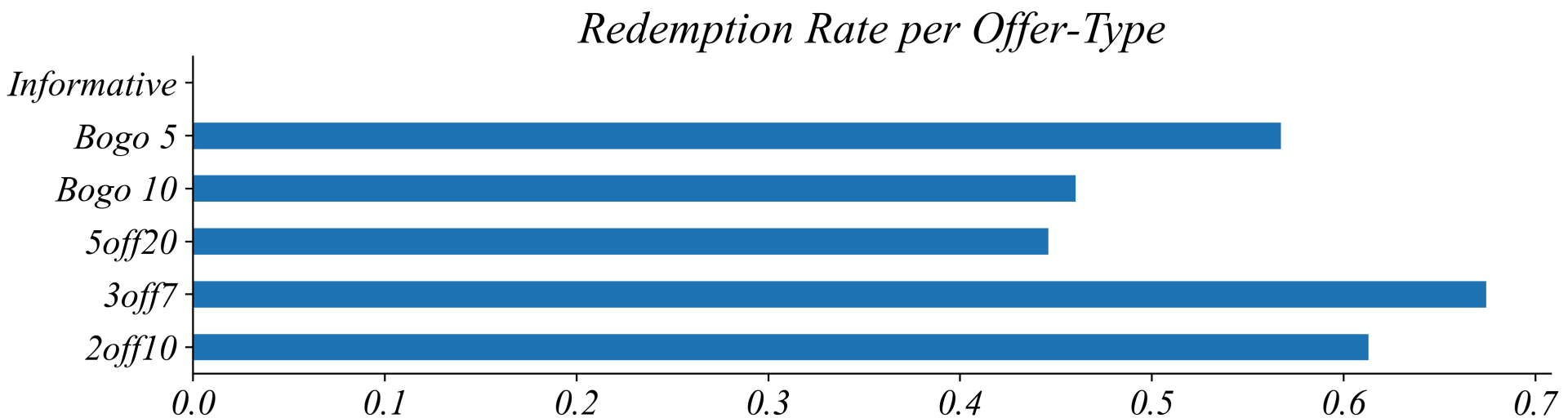
What's the average revenue per offer-type?

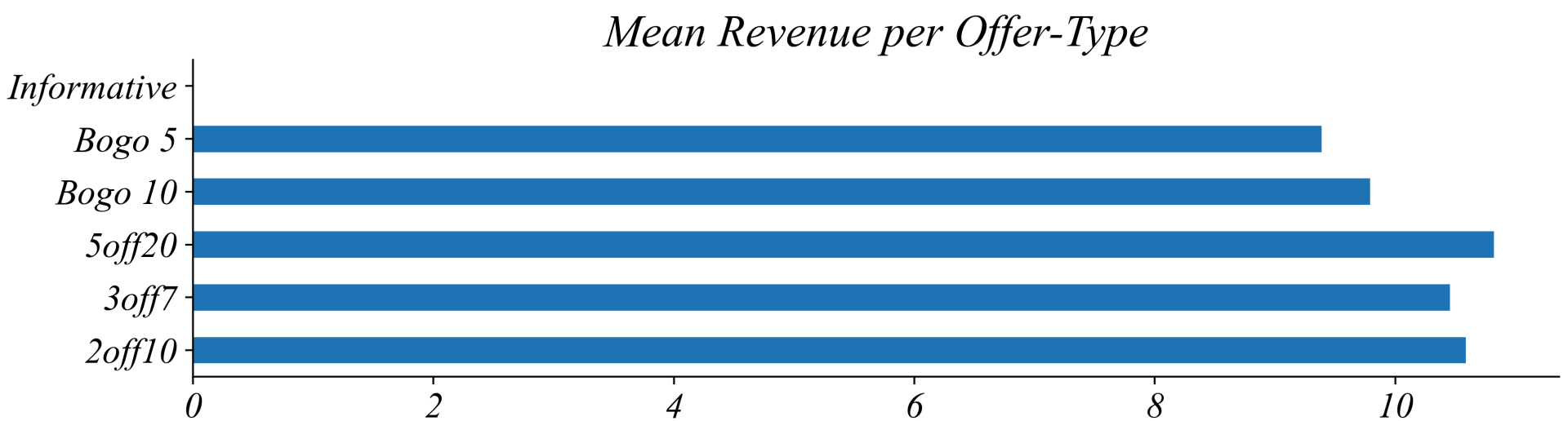
Lets find how much revenue we can expect after sending each offer-type.



Starbucks Offers: Effectiveness

Which offers are truly most effective?





Starbucks Offers: Conclusions

Q. Which promotional offers change behavior the most?

*1. The offer 3off7 has a high **redemption rate** but the discount may be costly.*

> people respond most to the 'best' deal

*2. The offer 5off20 has the highest **revenue** but a lower **redemption rate**.*

> people will spend the most when its required to redeam the offer

*3. The offer 2off10 is a more modest discount and is second in on **both metrics**.*

Part 1.7 | Grouping

Summary

- ***Filtering*** *out irrelevant rows before grouping*
- ***Group*** *by relevant columns to quickly summarize data*
- ***Aggregate*** *using sum, count, mean, max, etc.*

Exercise 1.7 | Starbucks Offers

Q. Which promotional offers change behavior the most?