



*Enabling  
Arbitrage  
via*



StubHub™

1. Hi, I'm Wade Webster, and this is my capstone project, "Enabling Arbitrage via StubHub."
2. So, I know what you're thinking, <next>

# What is *arbitrage*?

1. Just what is arbitrage?? <next>

***Arbitrage*** is the simultaneous purchase and sale of an asset to profit from a difference in the price.

1. Arbitrage is the simultaneous purchase and sale of an asset to profit from a difference in the price. <next>

It is a trade that profits by exploiting the price differences of identical or similar financial instruments on different markets or in different forms.

1. It is a trade that profits by exploiting the price differences of identical or similar financial instruments on different markets or in different forms. <next>

*Arbitrage* is  
totally not scalping.

1. Arbitrage is totally not scalping! <next>

(But, it *kinda* might be.)  
More on that later.

1. But, it kind of might be.
2. More on that later. <next>

# First Things Come First: Shortcomings

1. So, first things first, let's talk about the shortcomings of this project.
2. First up is the setup.
3. My first attempt was to setup a recurring Python script on Amazon Web Services.
4. The script is activated every thirty minutes and transforms a JSON query into a more manageable .CSV file.
5. But, any IP address from Amazon Web Services is blocked by StubHub.
6. Fortunately, I was able to migrate my original work to Google Cloud Compute.
7. Unfortunately, at the time I did not code any error handling, and the API response has a nasty tendency to return malformed data. <next>

# And also: Assumptions

1. Up next is assumptions.
2. I have to assume that sellers are the original ticket purchasers paying face value.
  1. Also, only one ticket is sold per listing.
  2. Why only one ticket? Well, I have no count for final sales, and I can only conclude completed transactions from when a listing drops stops updating.
    1. The data StubHub provides is characteristic of available listings, not for transaction clearances.
  3. I use these assumptions to calculate profits. <next>

# Further more: Comprehensiveness

1. And finally, comprehensiveness.
2. This research is built around observing two high volume listing events:
  1. Fireworks Night for the San Francisco Giants on May 26, 2017
  2. John Legend concert in Berkeley, CA on May 27, 2017
  3. Each data set came in at about 350,000 rows of data in a 90 MB CSV file
3. Also, my project was conceived with an alternative intent compared to the final product.
4. I intend to further develop this... to further my understanding and knowledge of this market, or for personal monetary gains.
5. So, let's talk about development. <next>

# Project Development Cycle

1. Altruism
2. Personal Goldmine
3. Product

1. This project has gone through three distinct phases: Altruism, Personal Goldmine and Product.
2. So, phase one, <next>

# **1. Altruism**

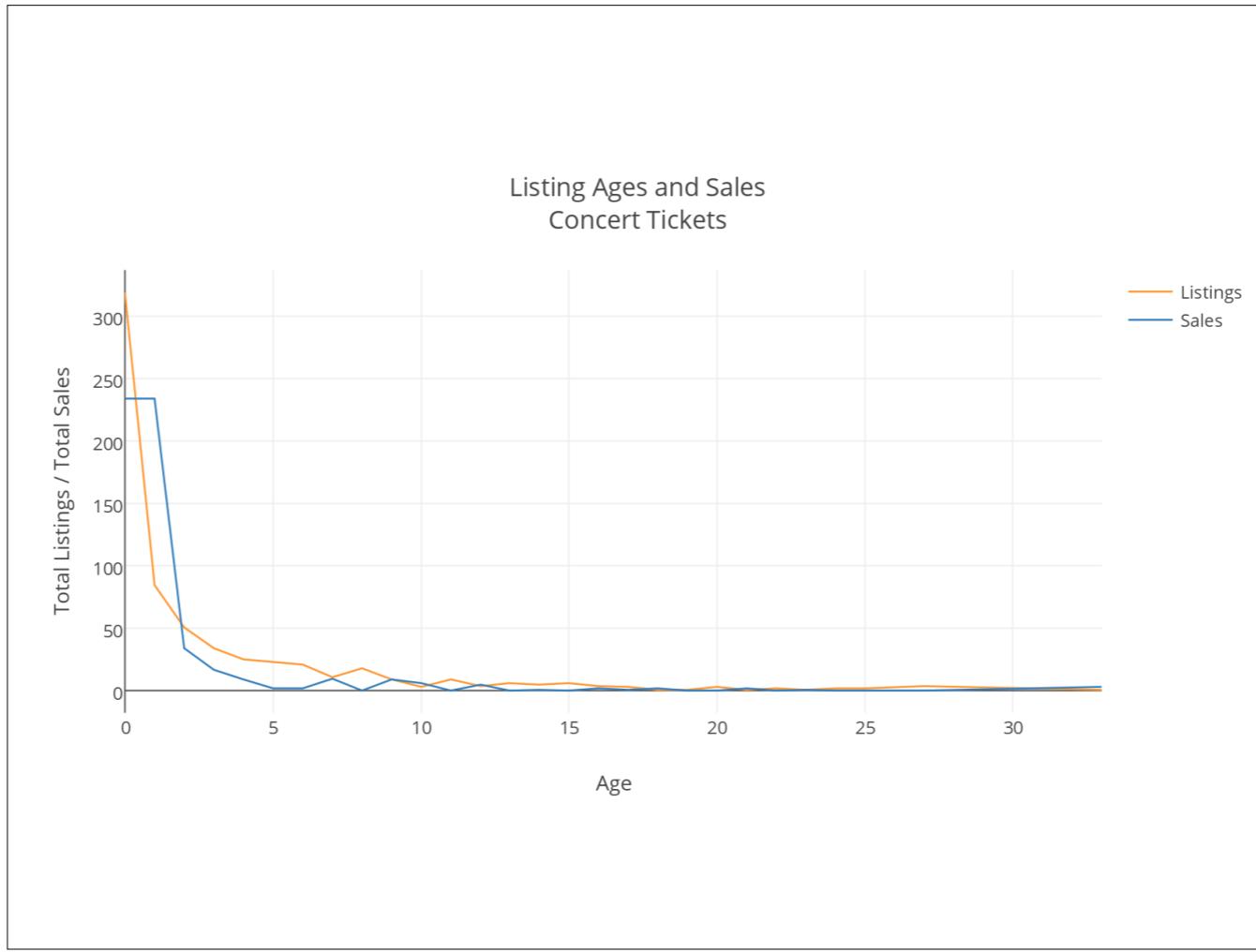
**What comprises a profitable listing?**

1. Altruism.
2. What makes for a profitable listing? <next>

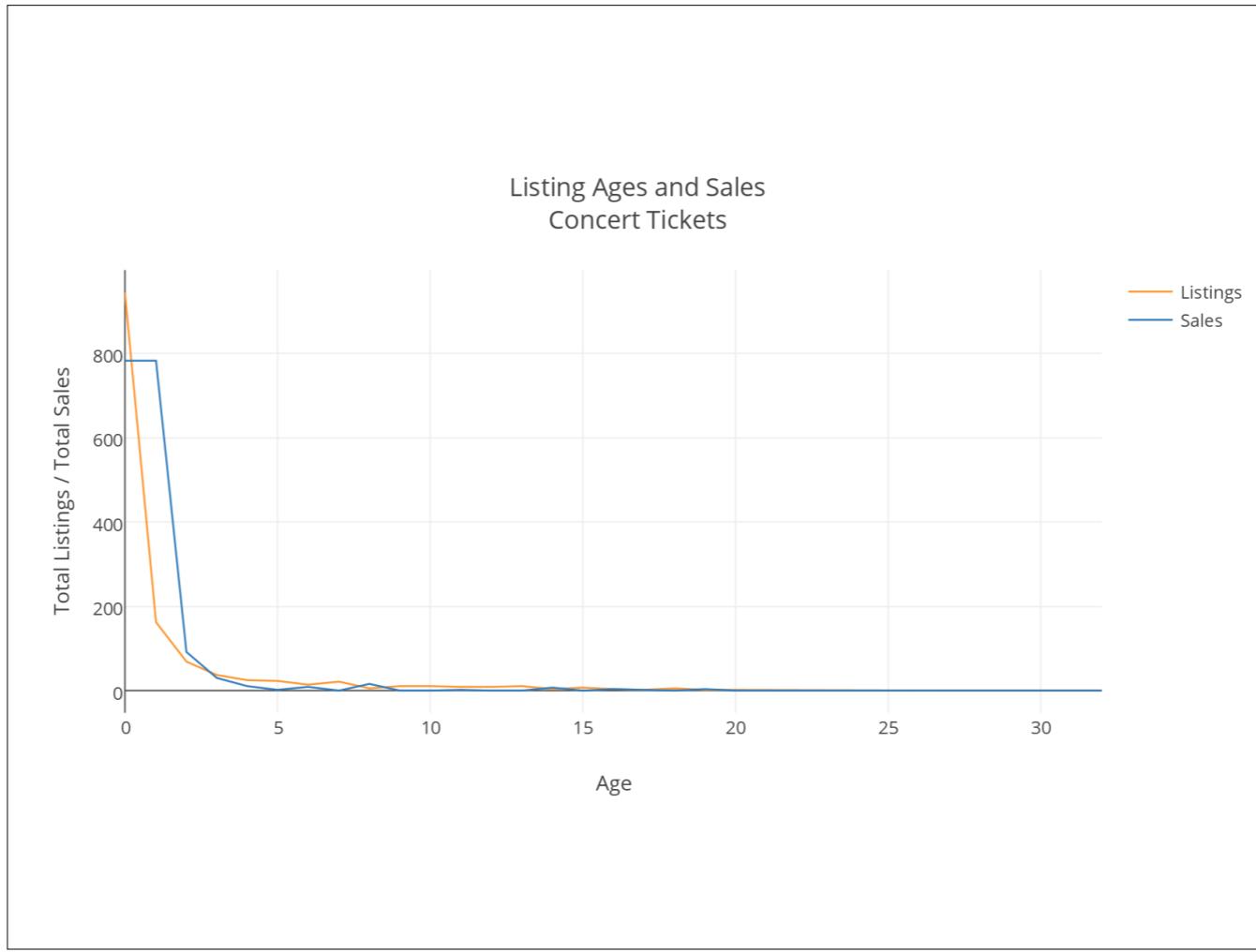
# Phase 1: Altruism

- Circumventing a competitive landscape and providing a social good
- Monetizing development through content marketing
- Advancing my reputation with a cool product

1. My original idea was a platform where users can upload their ticket information and receive recommendations for optimizing a StubHub listing.
2. Currently, there are paid services that do this, such as TicketFire or TicketFlipping, and working around them would personally be, I don't know, fulfilling, for lack of a better word?
3. So, let's talk about what makes a successful StubHub sale.
4. First up is ticket age. <next>



1. Here, we see the total volume of concert ticket listings against sold tickets over 30 days.
2. The majority of ticket transactions are within the first three days. <next>



1. This behavior is similarly represented within the sports event.
2. Any listing with a lifespan of greater than three days is otherwise stagnant.
3. Onward to...<next>

# A Tale of Two Targets

## Profit and Value

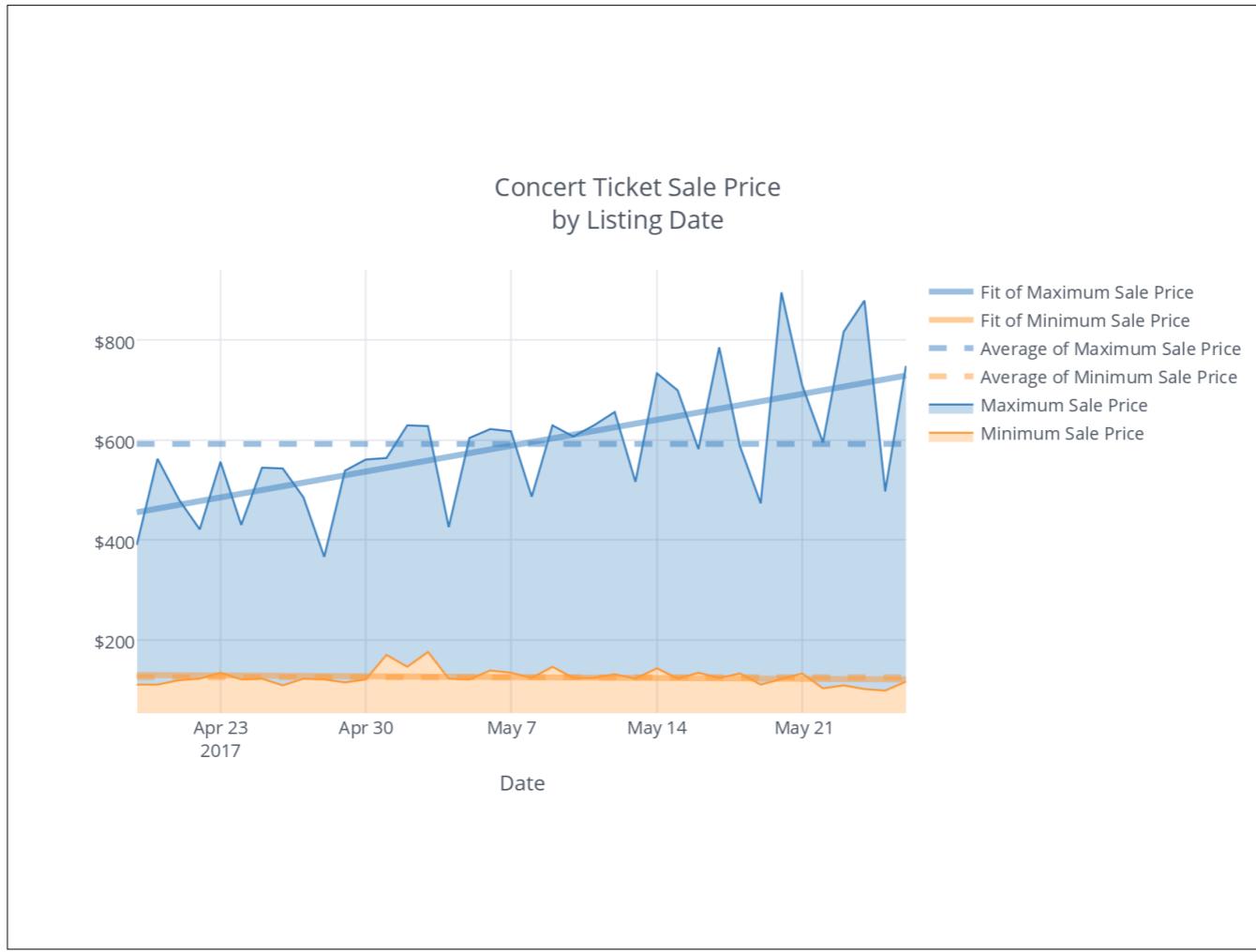
1. Our two targets: Profit and Value.
2. To successfully flip tickets, I need to accurately predict which tickets are both profitable for resale, and the conditions that represent the most value to a potential buyer.
3. These targets are engineered manually, and they are calculated slightly different...<next>

# Sale Price vs. Sold Price

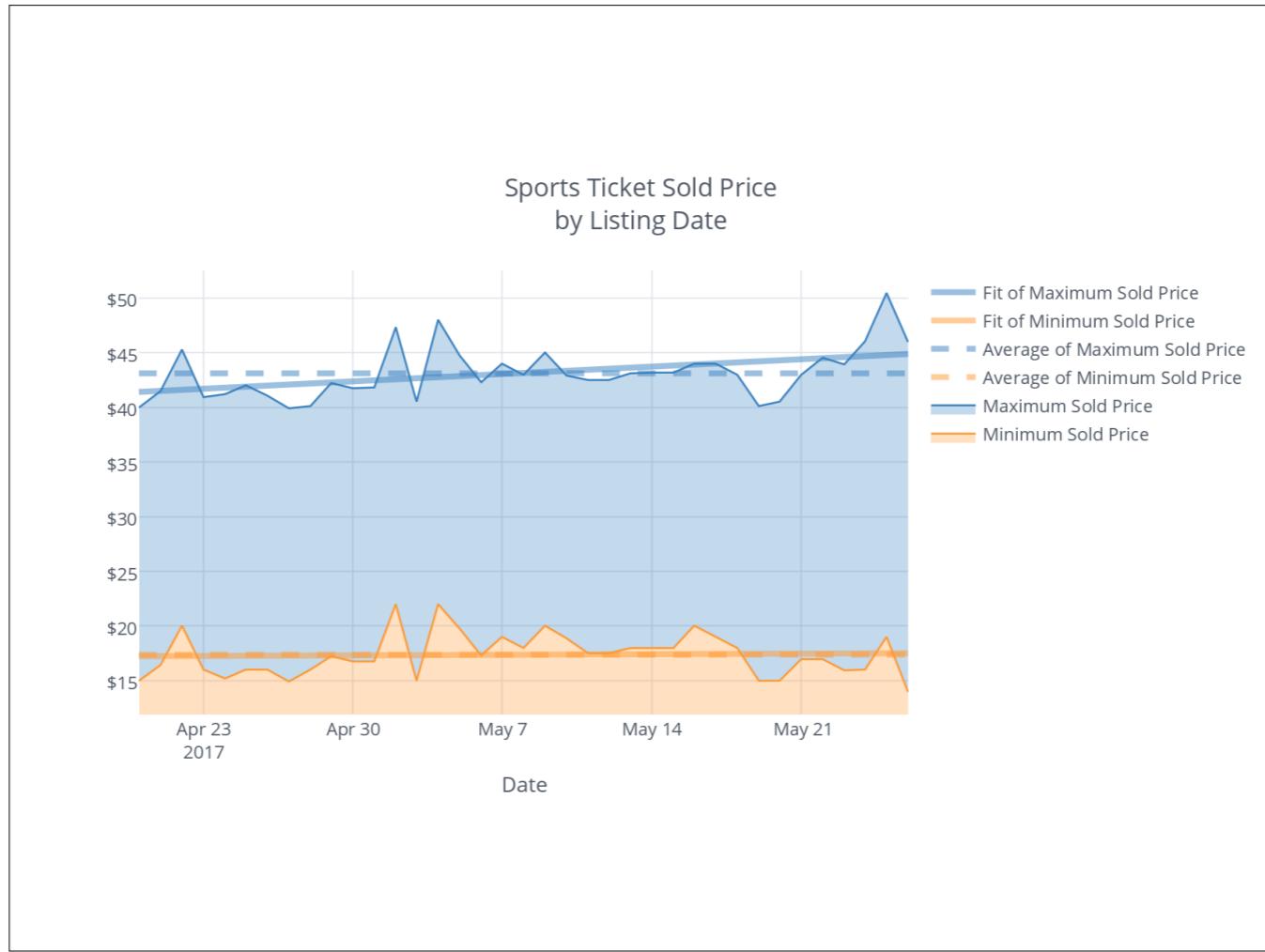
Profit = Sale Price - Average Face Value - (Sale Price \* .1)

$$\text{Value} = \frac{\text{Sold Price} - \text{Average Face Value}}{\text{Average Face Value}}$$

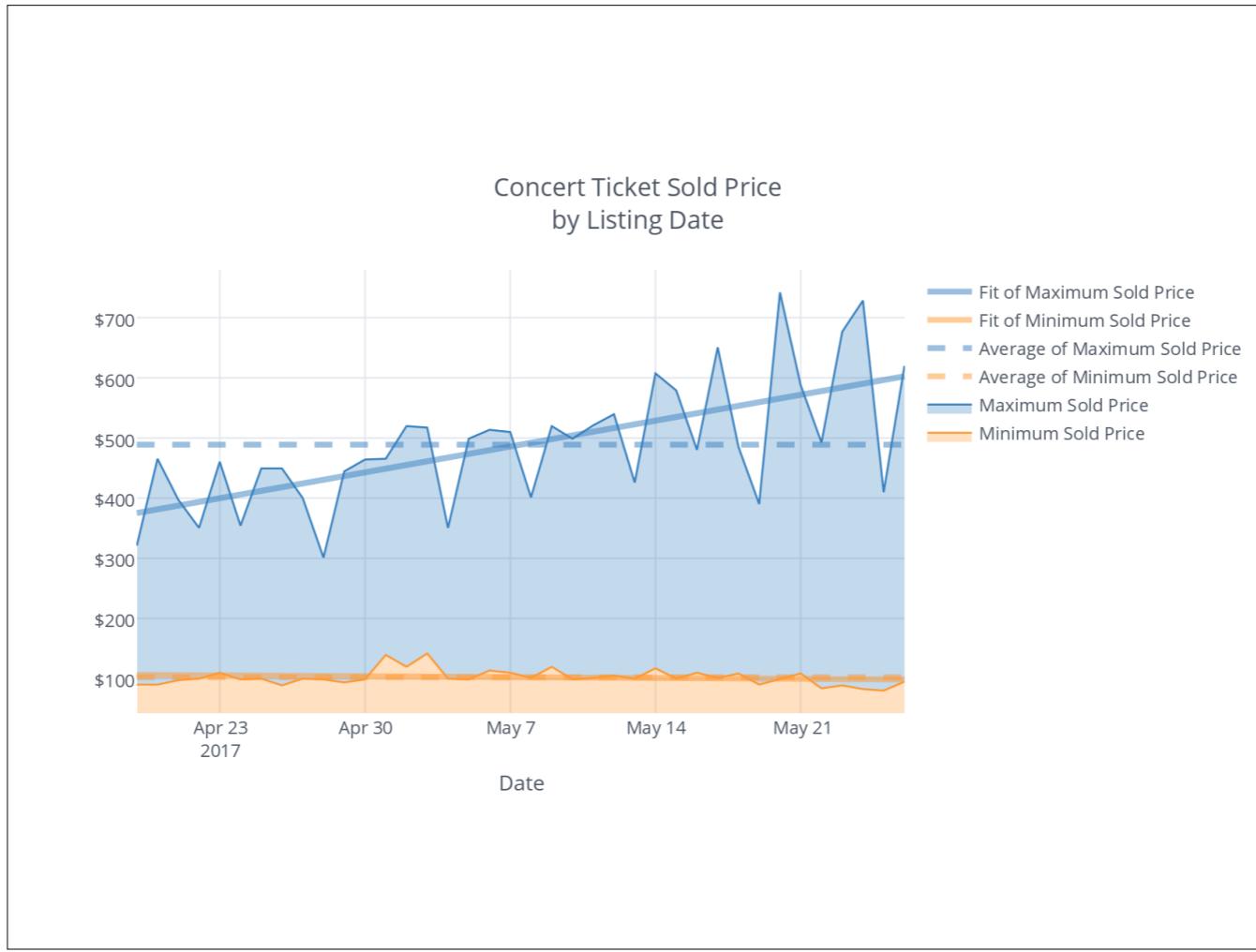
1. I've defined Profit to the seller as the Sale Price minus the face value of the ticket, which is a sunk cost, and minus 10% of the Sale Price, which is the fee a seller pays StubHub at sale of the ticket.
2. Value is calculated as the Sold Price, which includes a dynamically calculated fee to the buyer, somewhere in the range of 10-20% of the purchase price, minus the face value of the ticket (the intrinsic value). This is essentially the markup.
3. I divide this by the average face value to calculate a ratio between the markup and the intrinsic value of the ticket.
4. This is a scalar centered around 0.
5. Now, let's discuss the behavior of sale pricing and sold pricing over the lifetime of these events. <next>



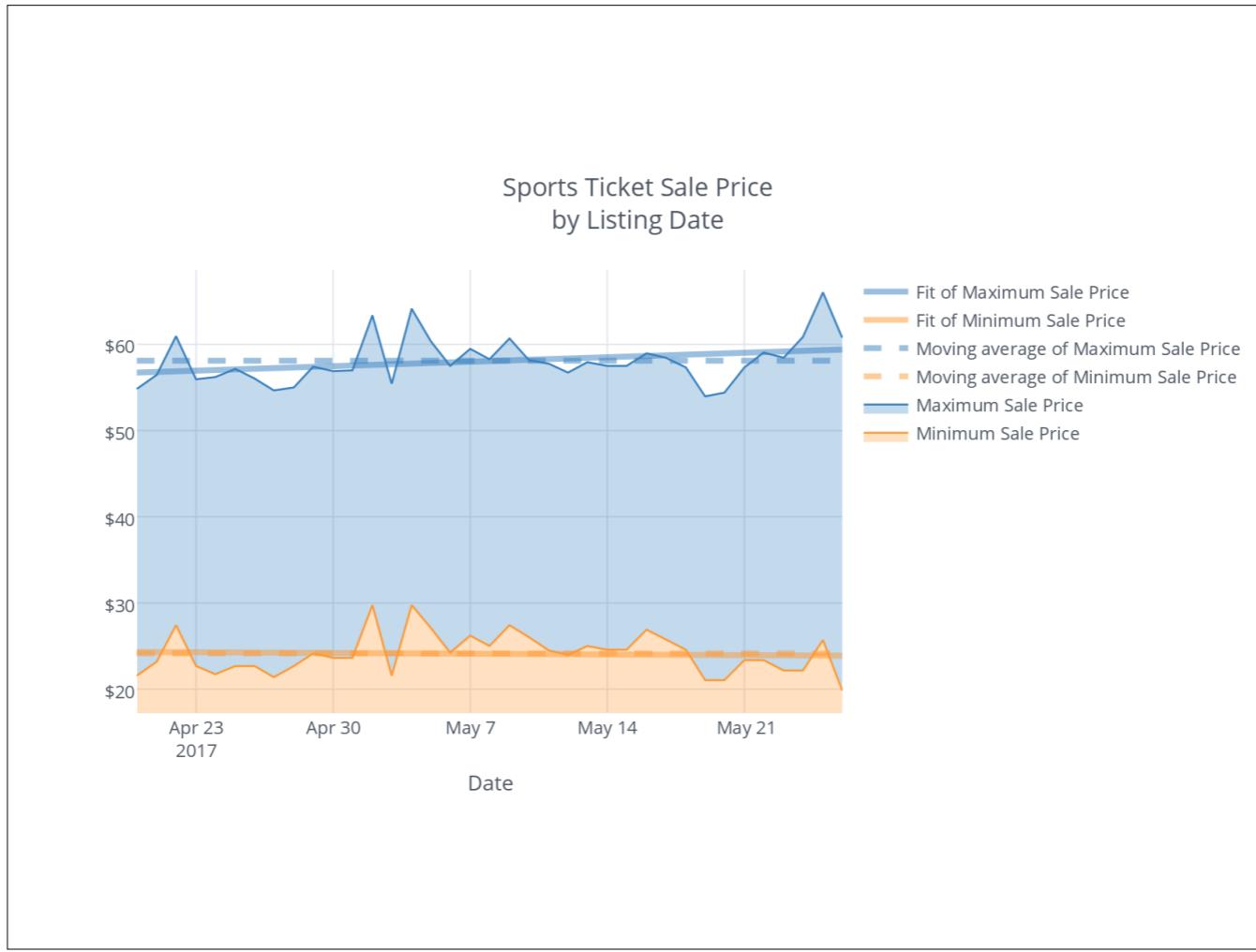
1. Here we have concert ticket sale prices.
2. An increased scarcity of premium seating drives average sale prices as the event approaches. <next>



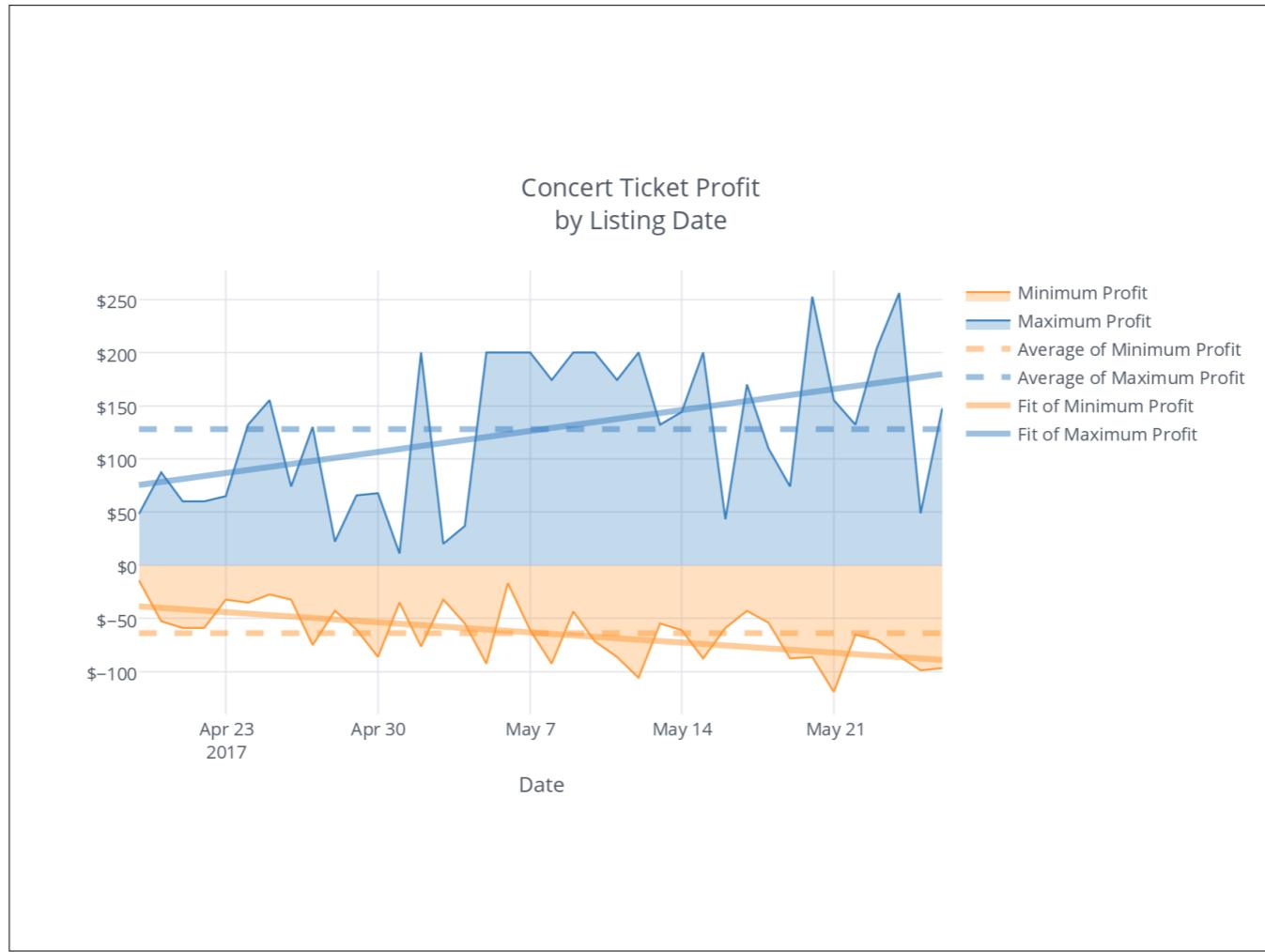
1. Whereas sports ticket remain relatively flat for the duration. <next>



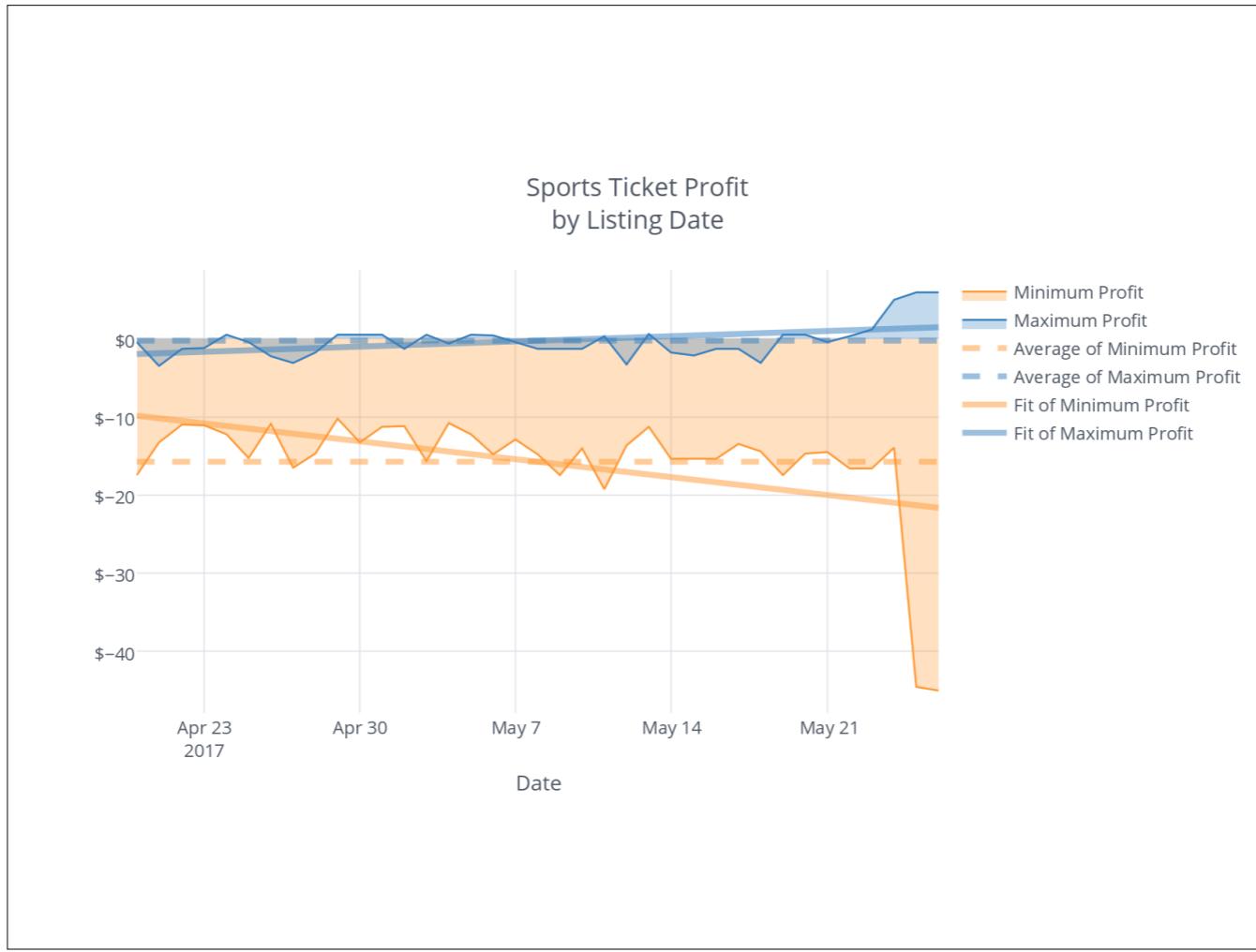
1. And here, the purchased price for concert tickets continues to grow as the event nears similar to the sale pricing.
2. As you will see...<next>



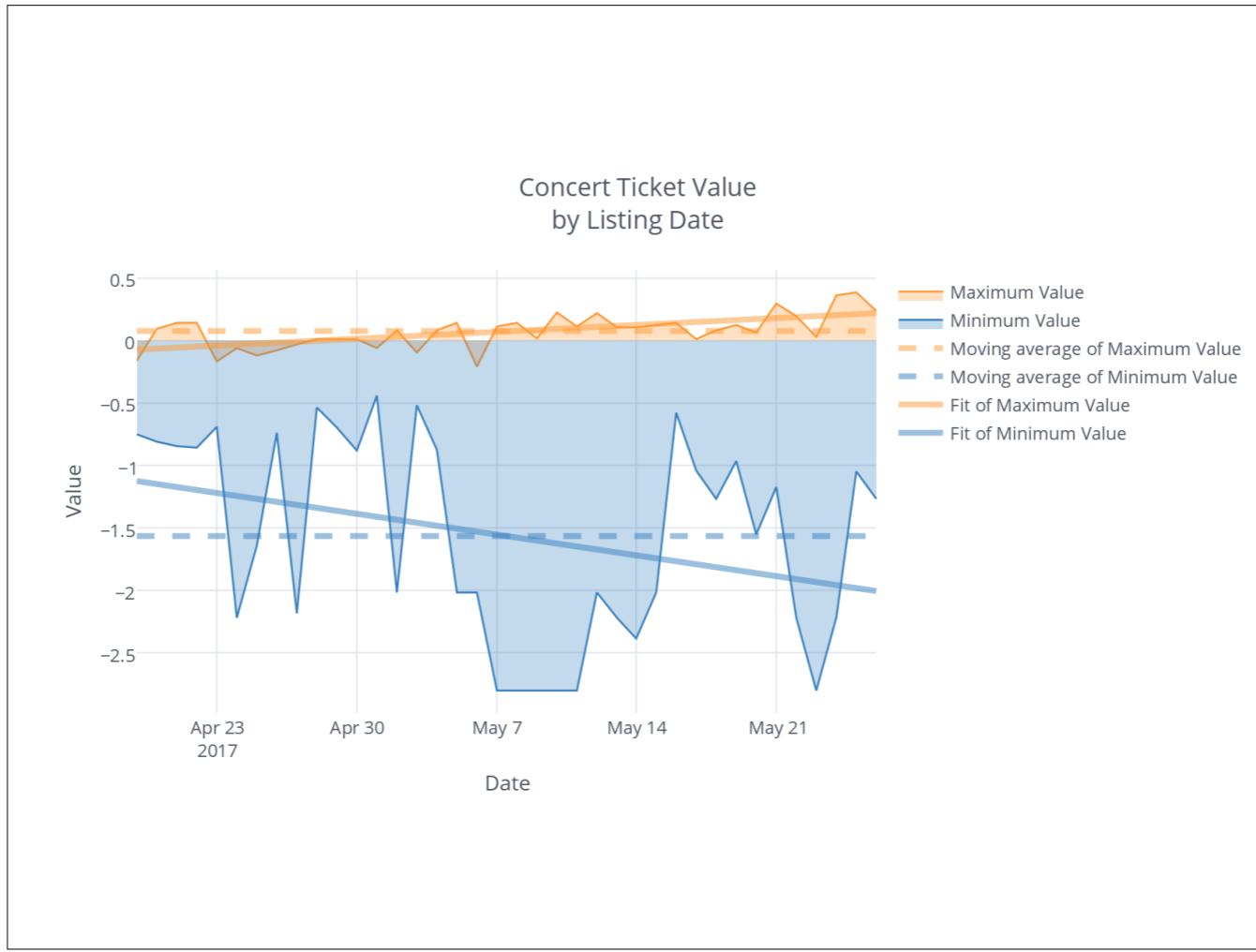
1. Alternatively, again the sports tickets remain relatively flat.
2. Next, let's check out profitability. <next>



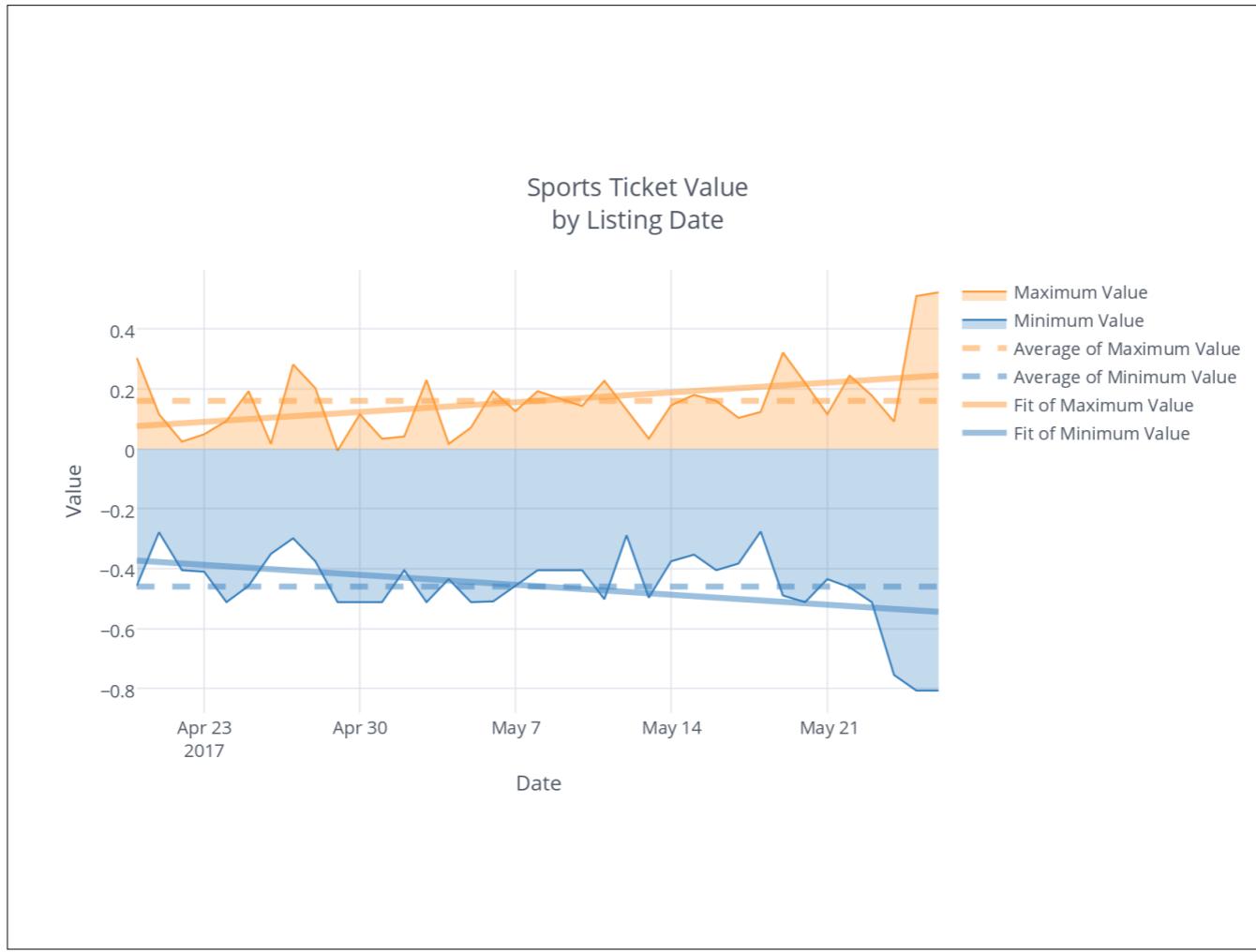
1. Overall profits for concert tickets grow steadily toward the event, but so does the negative range. <next>



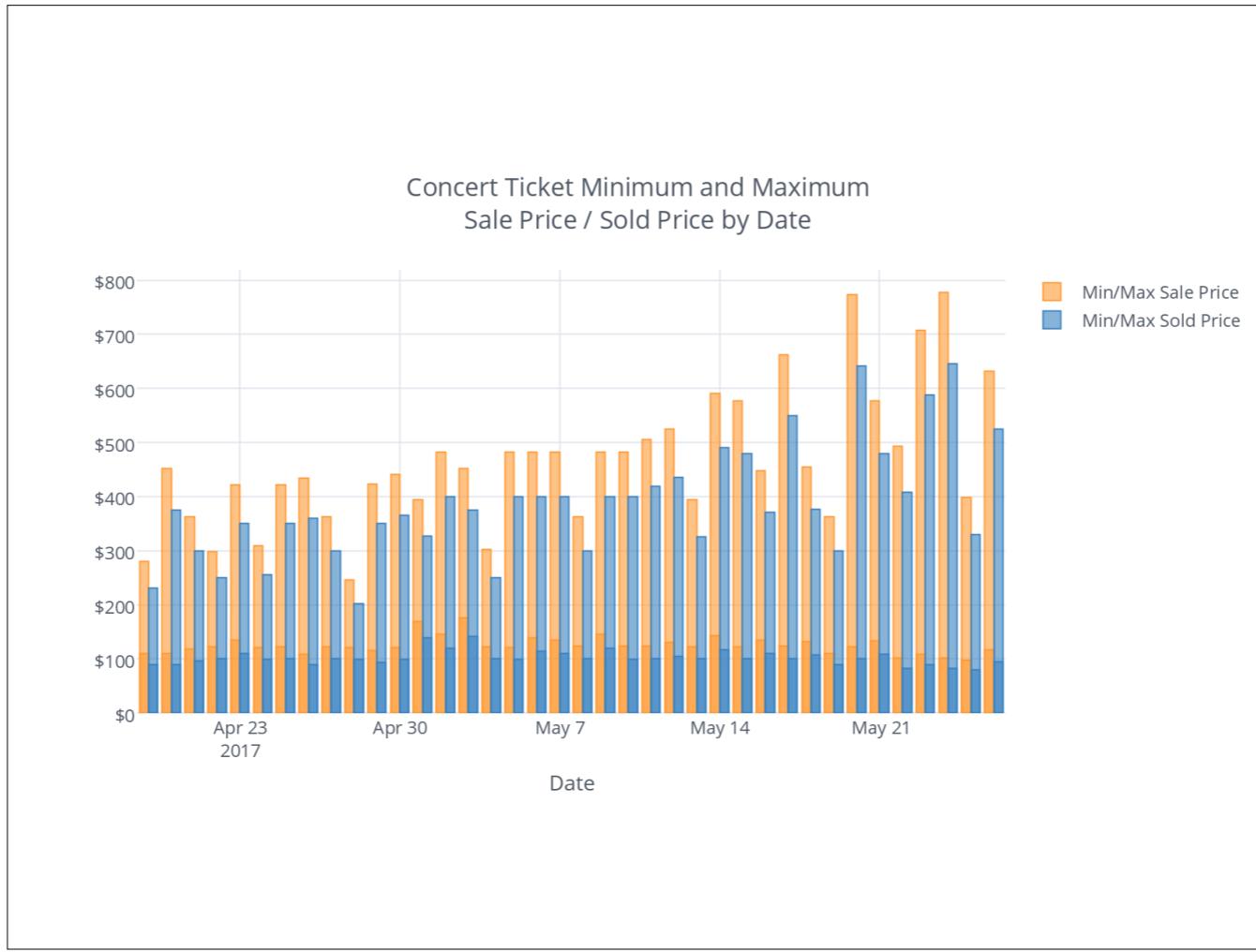
1. The majority of sports tickets are sold at a loss, and that loss only becomes more extreme toward the event.
2. Something to keep in mind is that, for the most part, ticket brokers, i.e. professional scalpers, are focused on events with large scale interest. The profitability for sports tickets are in playoff events, and the participants for those are largely unknown until close to the game.
3. The majority of sport ticket sellers are single sellers, that would be individual season pass holders. A person with a season pass, unable to attend, would likely take any money over nothing. Hence, the reduced profits. <next>



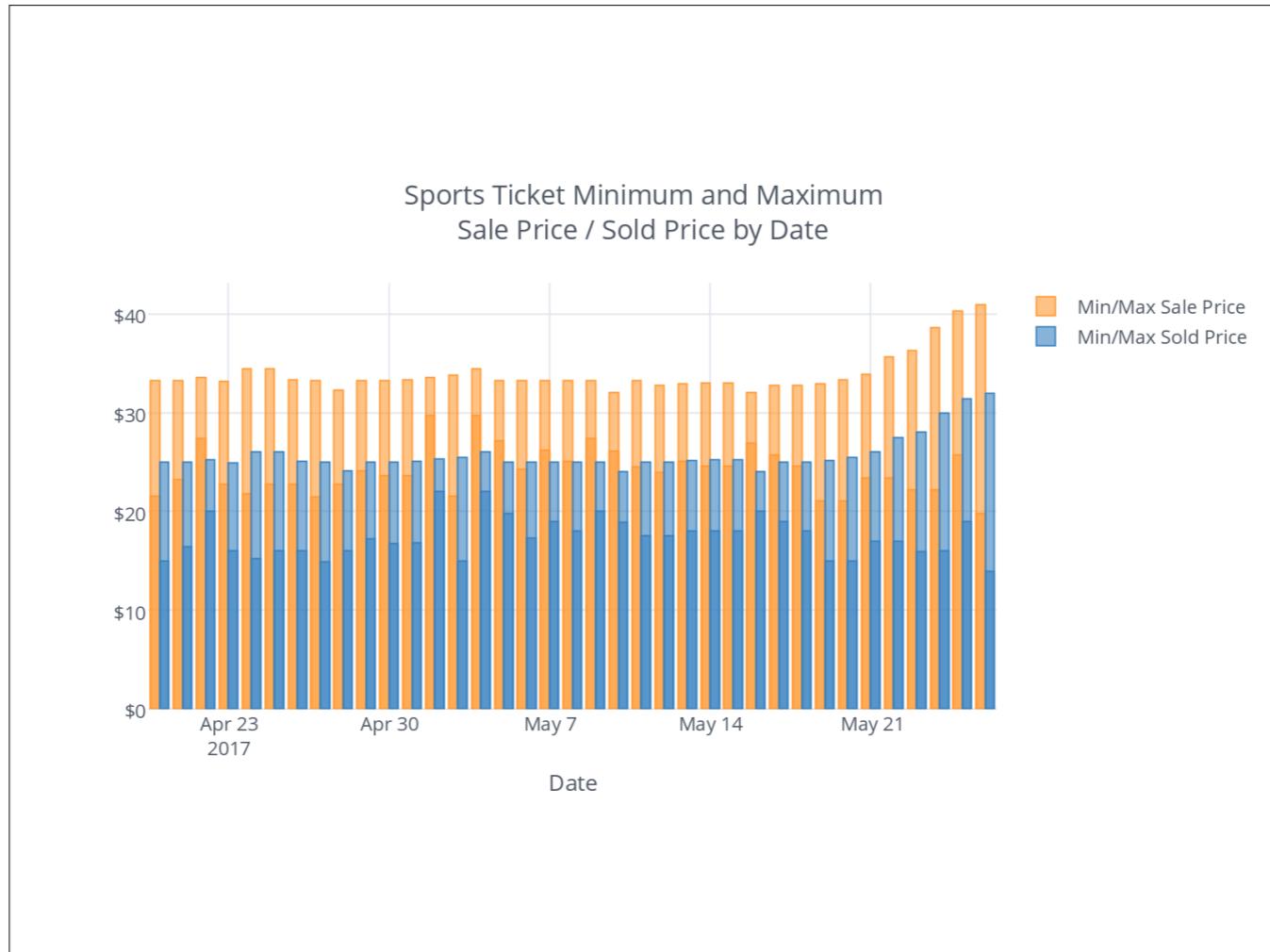
1. Now, onto value.
2. The maximum value for concert tickets grows slightly toward the day of the event, but the minimum value increases wildly.
3. There's a lot a speculative sellers out there.
4. Recall that value is calculated as a scalar centered around 0! <next>



1. The sports tickets are much more centered around 0, with increasing amounts of value toward the day of the event. <next>



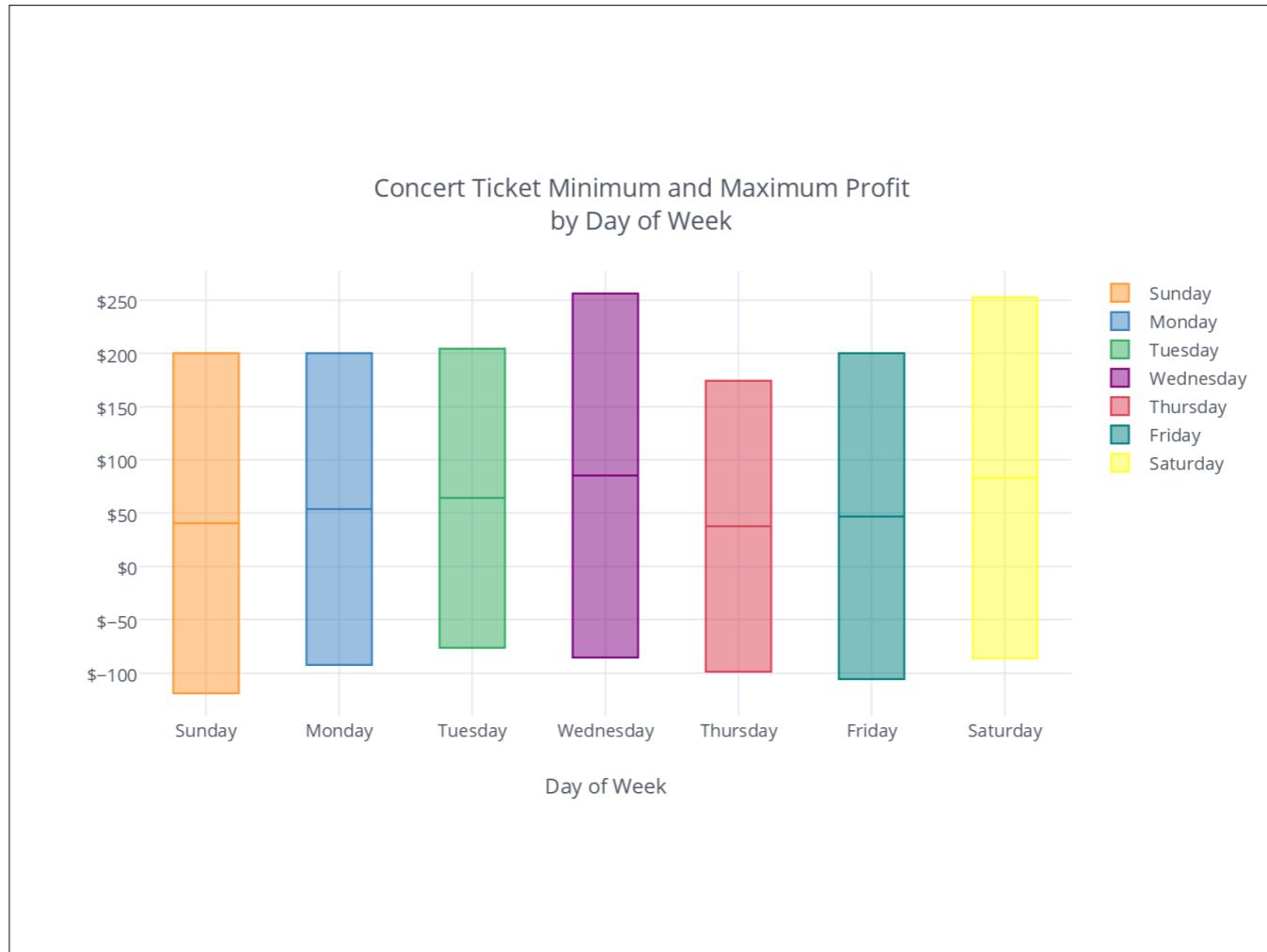
1. Here is the trend for all ticket pricing and sales up to the day of the concert.
2. Note the darker bars overlaid at the bottom.
3. These represent the minimum prices.
4. You'll see that there is limited fluctuation. <next>



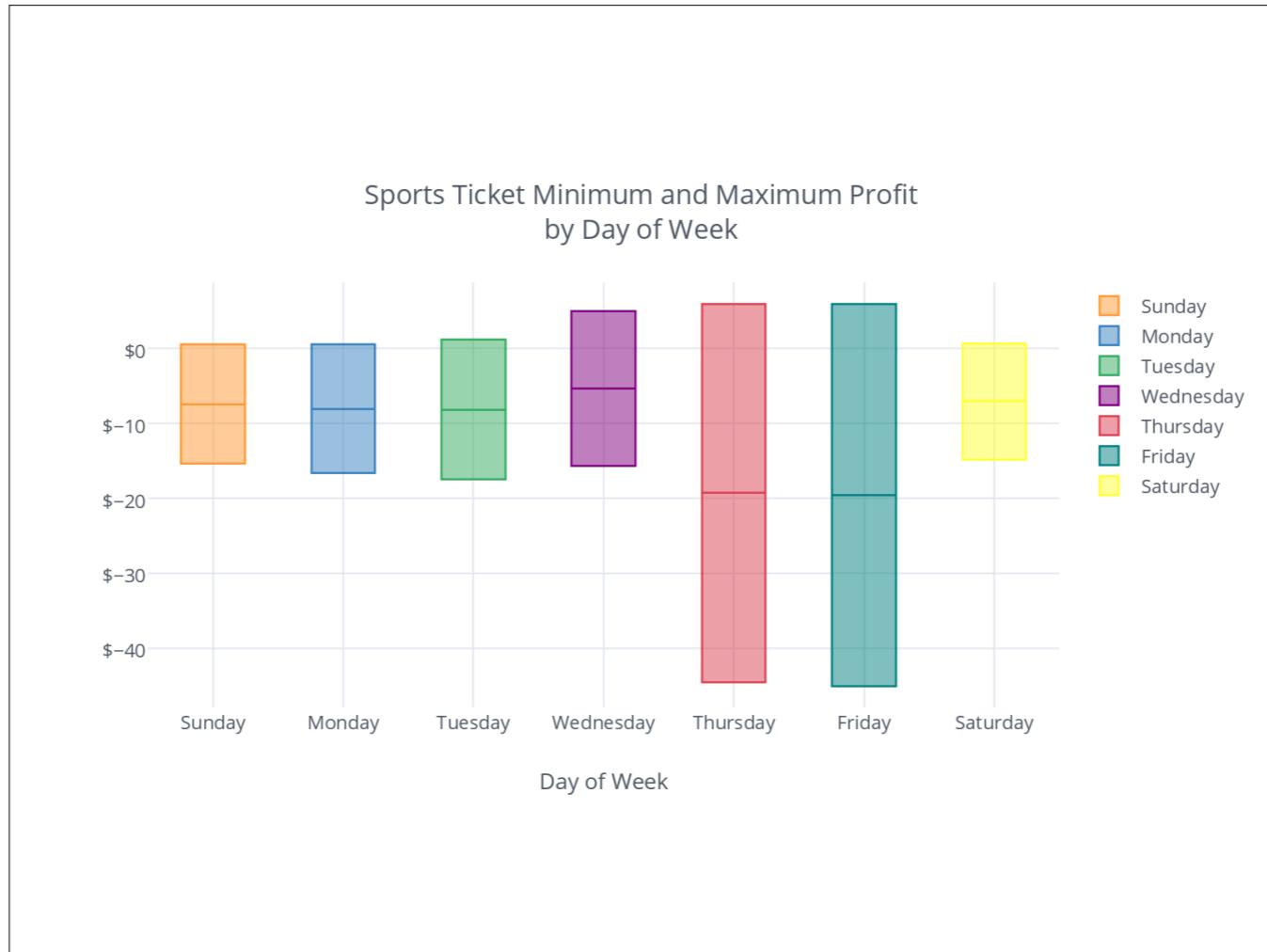
1. Whereas the transacted dollar amounts for the sports event only grows slightly within the final week.
2. The differences between minimums and maximums is much tighter than with the concert tickets, and the overall growth by the day of the event is only approximately \$6. <next>

# Maybe day or time of day?

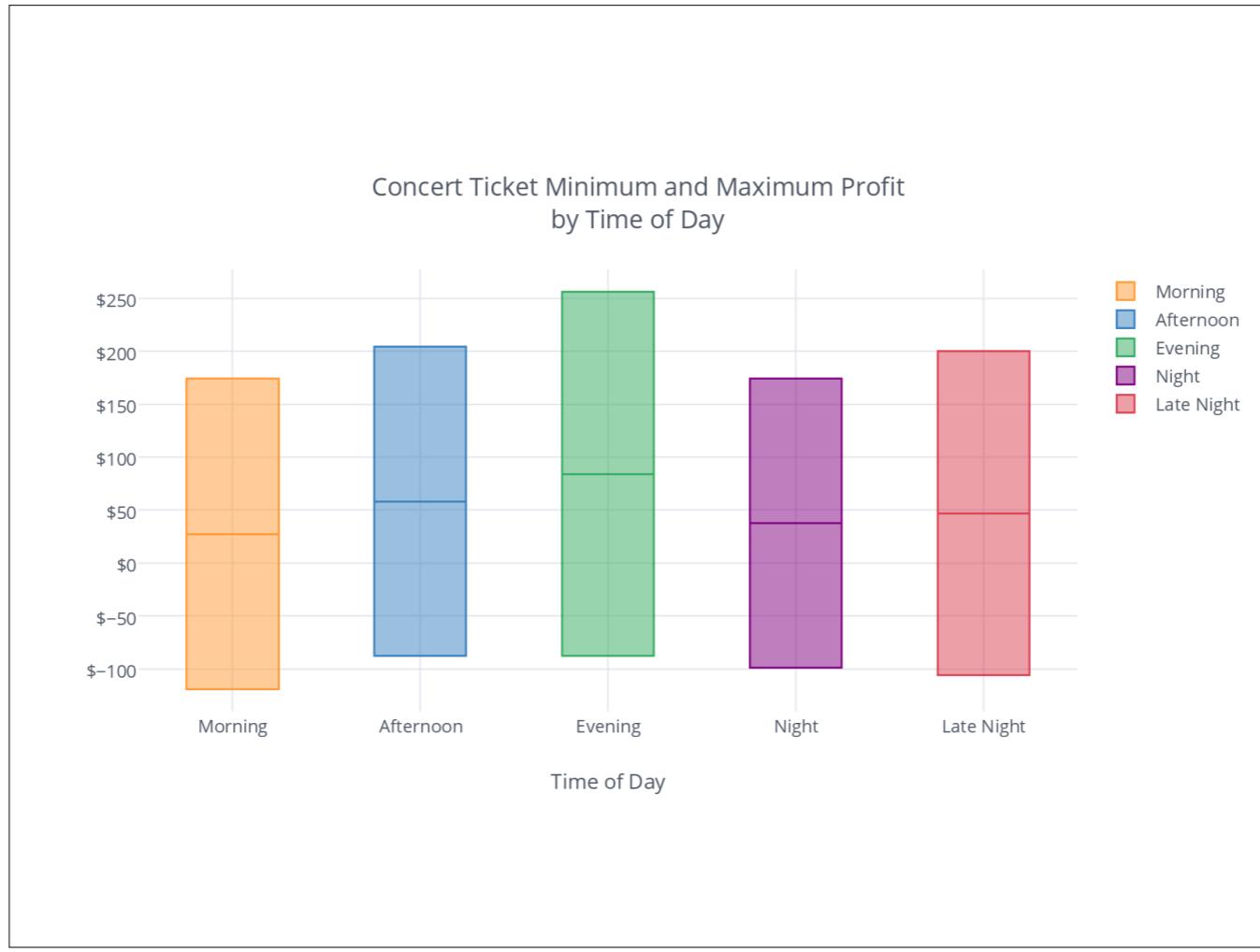
1. Next, I checked if there was any influence by the day or time of day a listing is launched. <next>



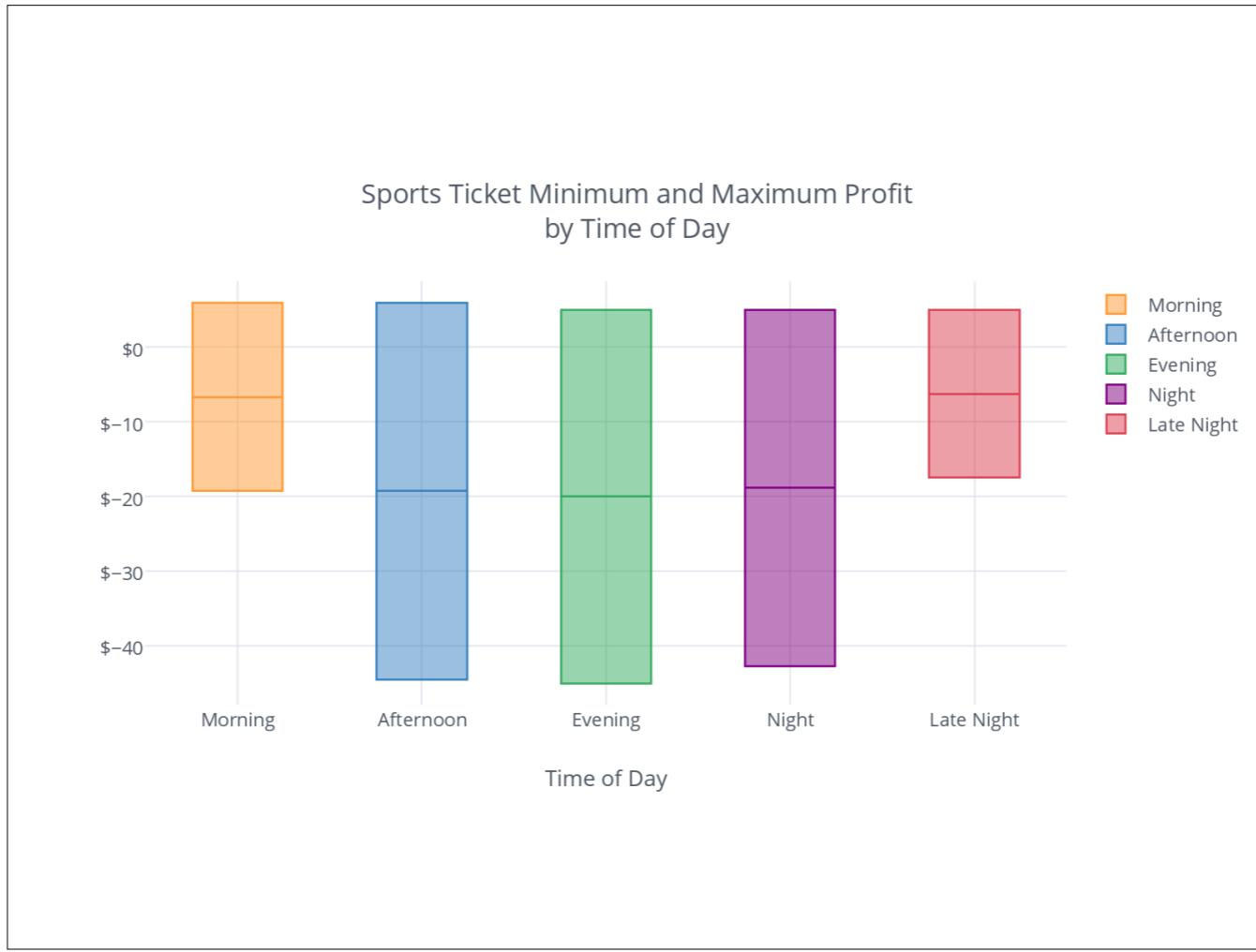
1. Wednesday and Saturday show an increased profitability for the concert. <next>



1. But, the sports event is relatively flat, save for the end of the work week when profitability tanks. <next>



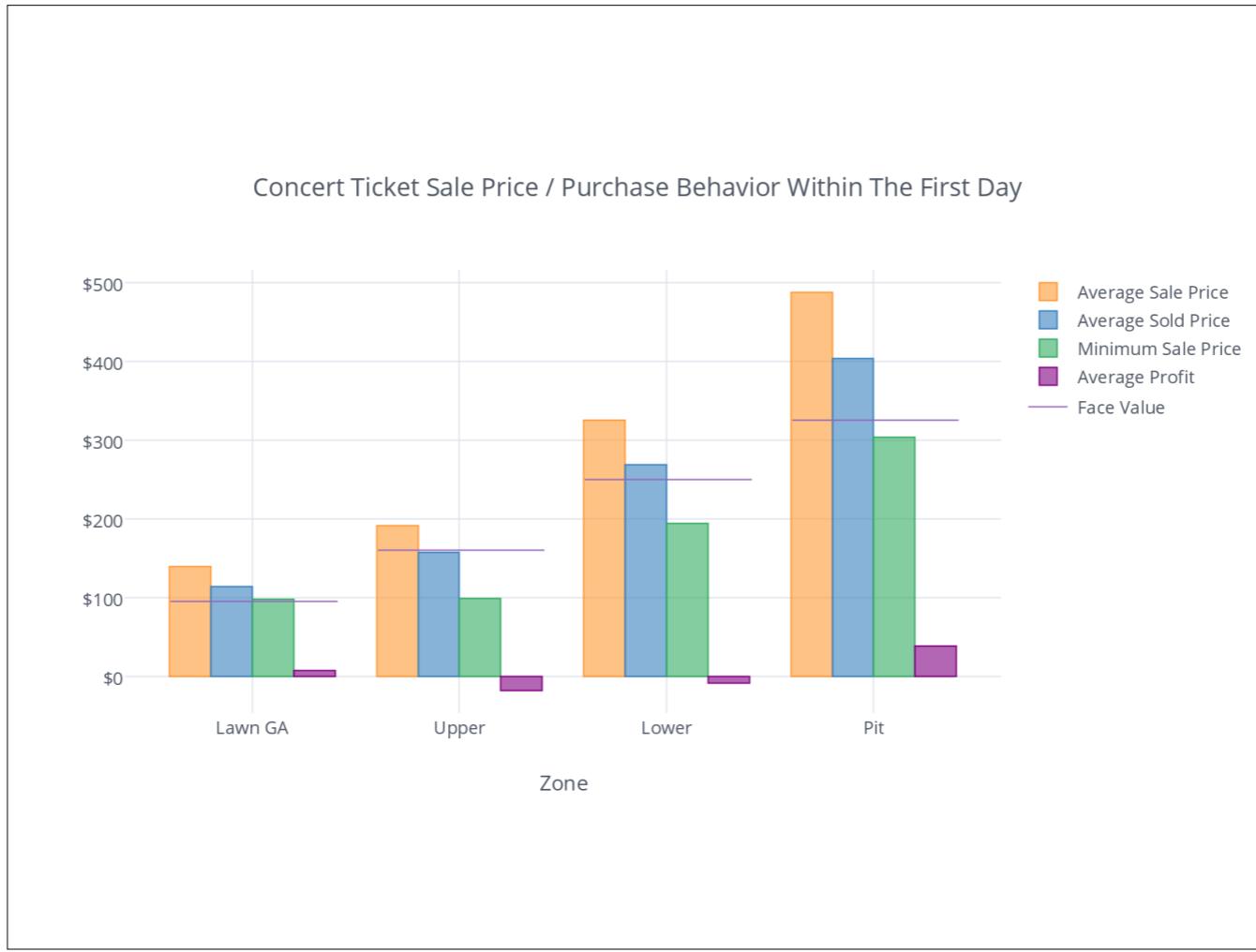
1. Next, I looked at the time of day.
2. The variance here is decent, and it seems to favor evenings, which I've binned as being from 5 PM to 9 PM. <next>



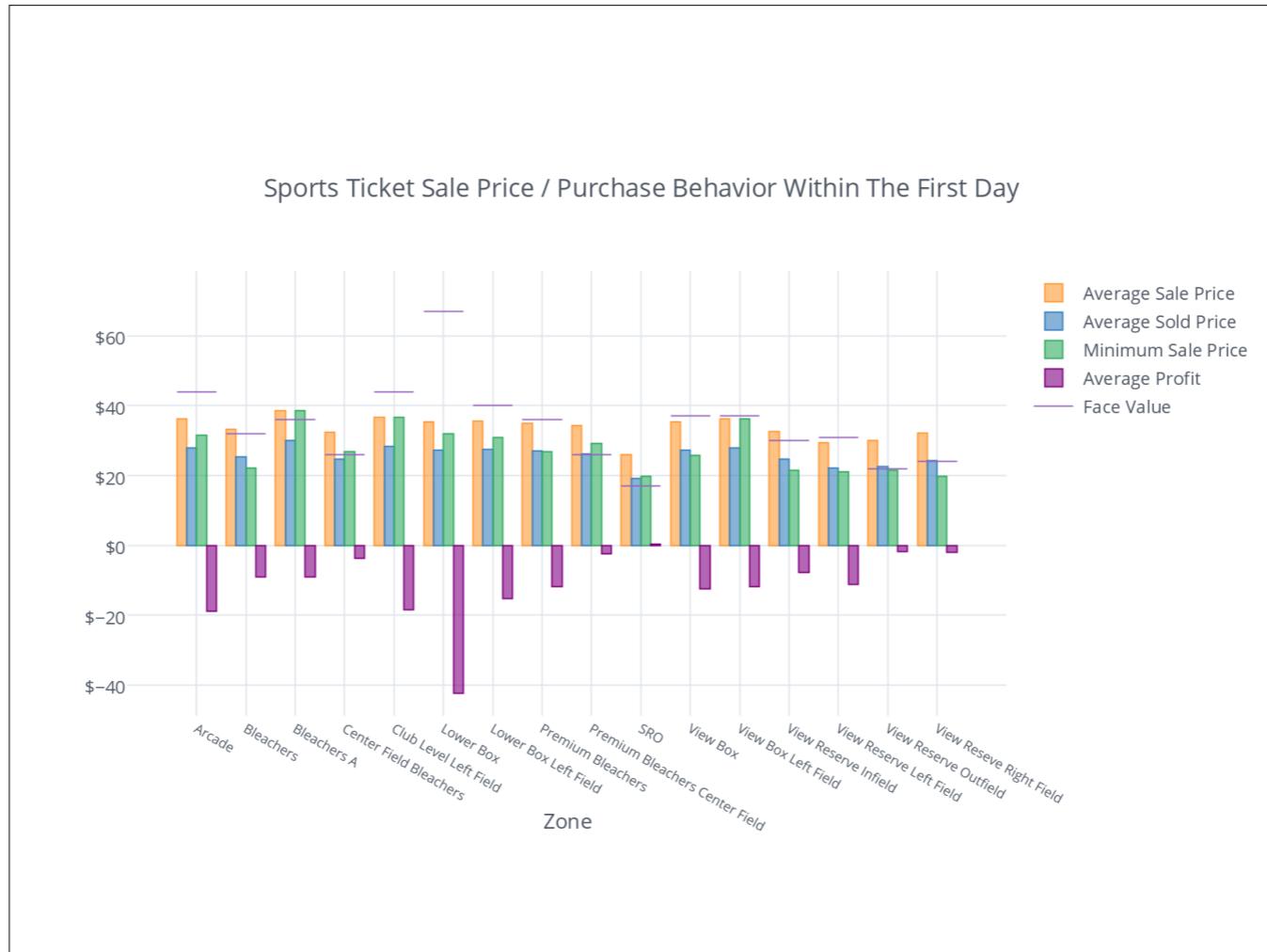
1. Strangely, for the sports event, profits are again flat, and the only scaling is downward. <next>

# Immediate Sales and the Long Tail

1. As we saw earlier, the majority of tickets are sold within the first three days.
2. For our resales to be most effective, I say the focus needs to be within the first twenty four hours.
3. Let's take a look...<next>



1. The zones for the concert tickets are ranked left to right from furthest to the stage to closest.
2. I've drawn a horizontal bar to represent the face value of a ticket for each zone.
3. As you can see, immediate profits for mid-range zones are on average negative, with profits existing only in the most inexpensive and expensive zones. <next>



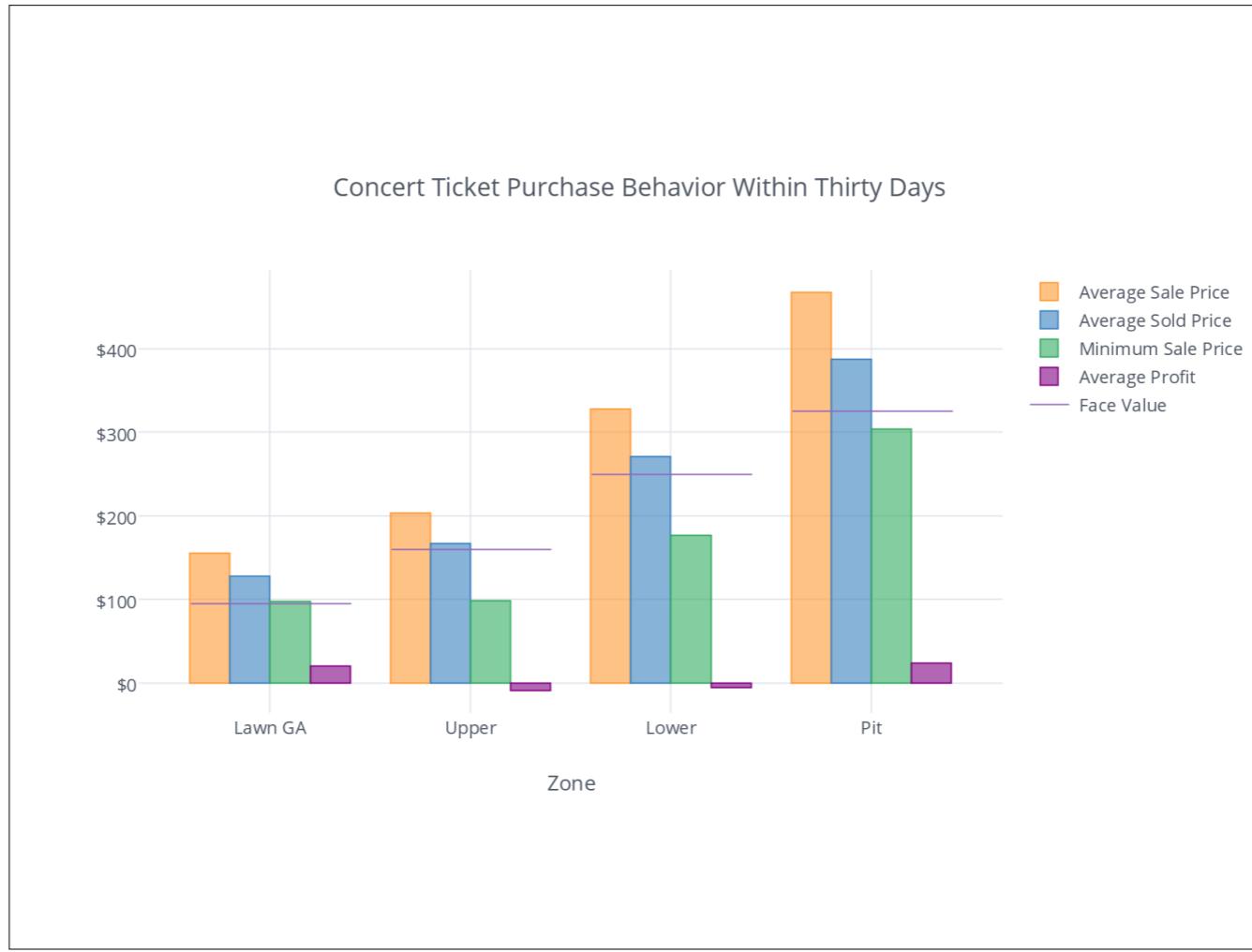
1. Sports event tickets are regularly sold at a loss, especially for midrange zones, such as 'Lower Box' seats. <next>



1. With that critical first three days in mind, let's expand just past that to five days.
2. Returns for mid-range zones remains negative, but the profitability of the cheapest zone has grown to almost match that of the premium zone. <next>



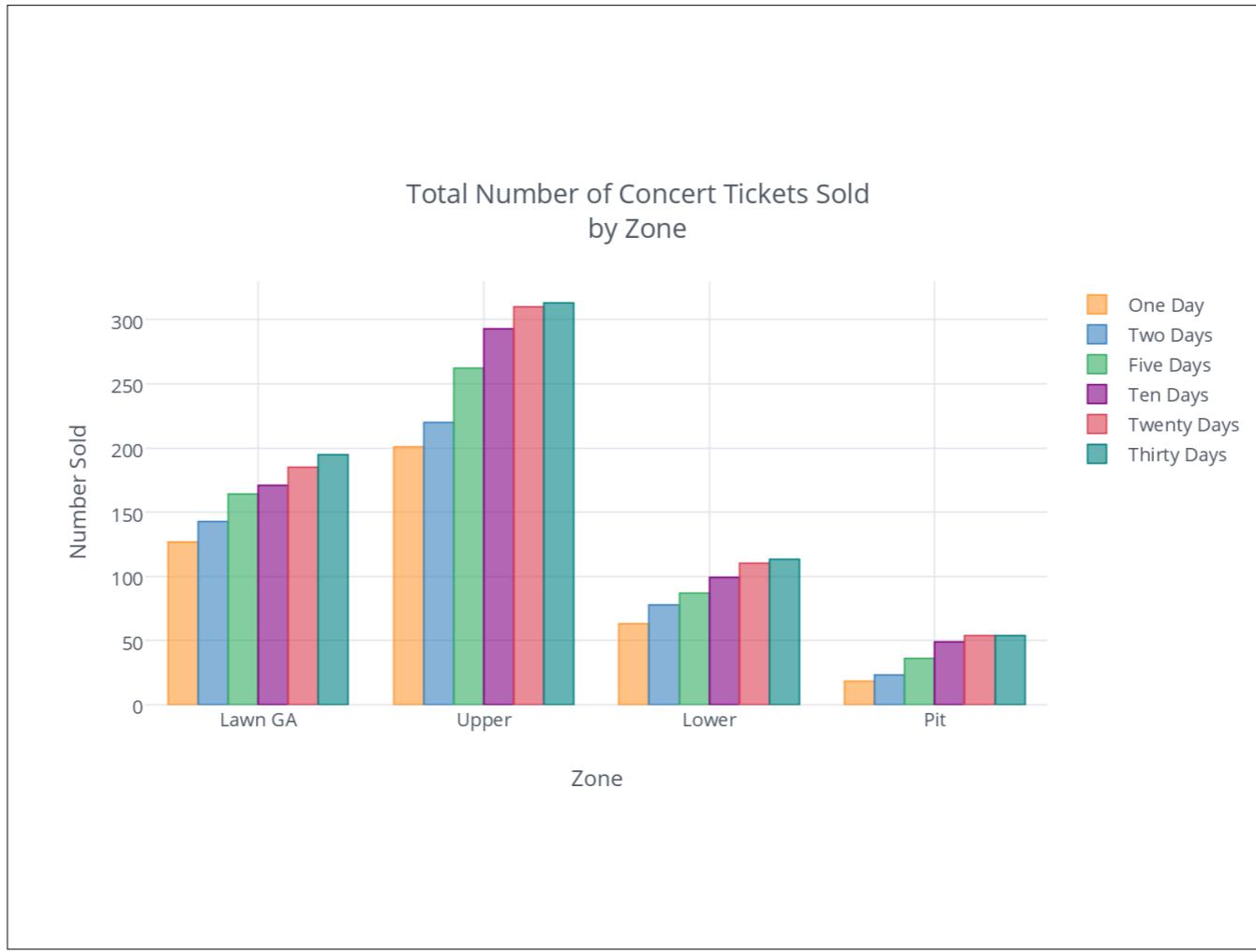
1. And the sports ticket market continues to stagnate with negative profits. <next>



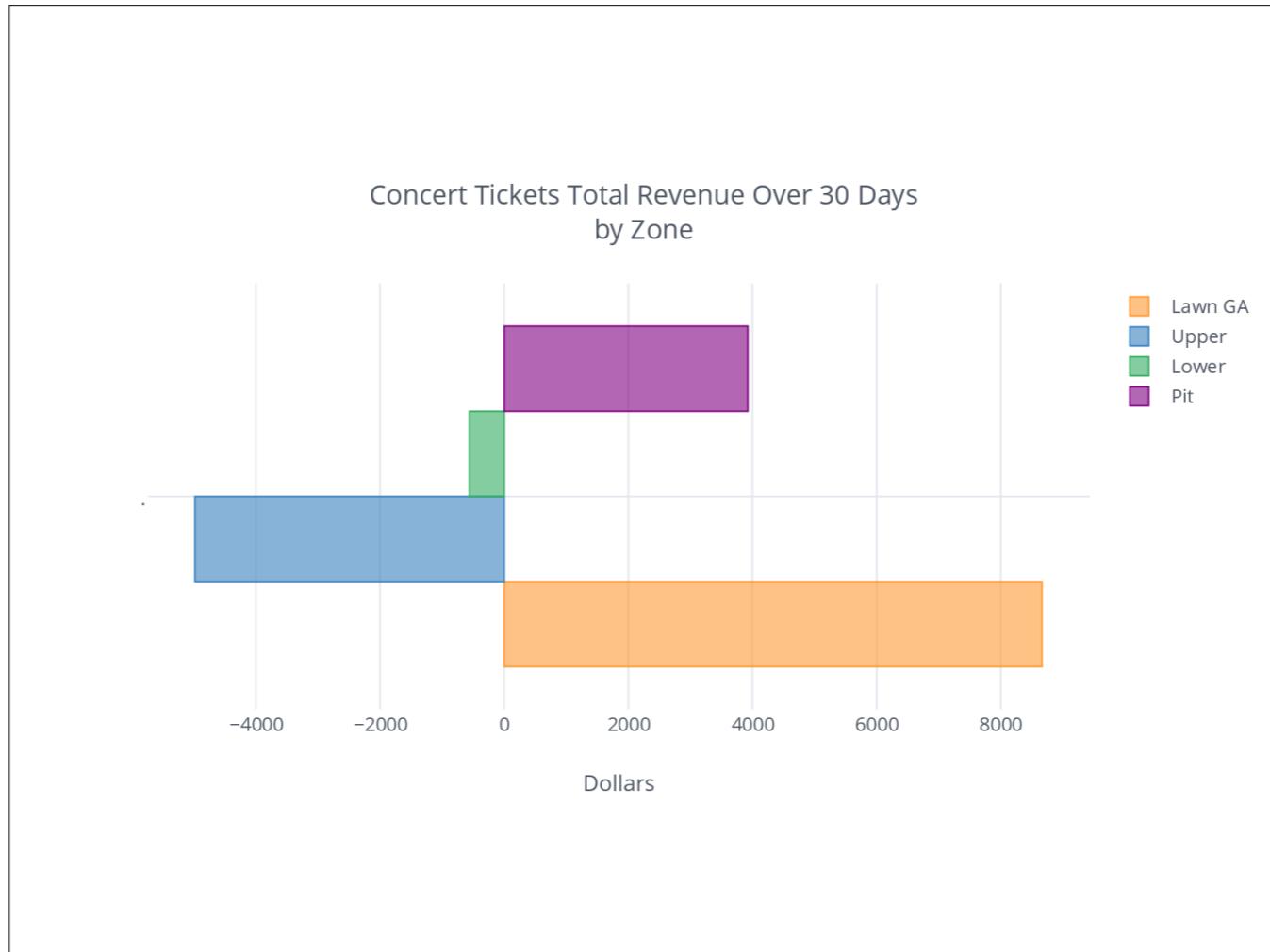
1. Here's a comprehensive window into an entire month of transactions.
2. At this point, the cheapest zone has almost surpassed the premium zone in profits. <next>



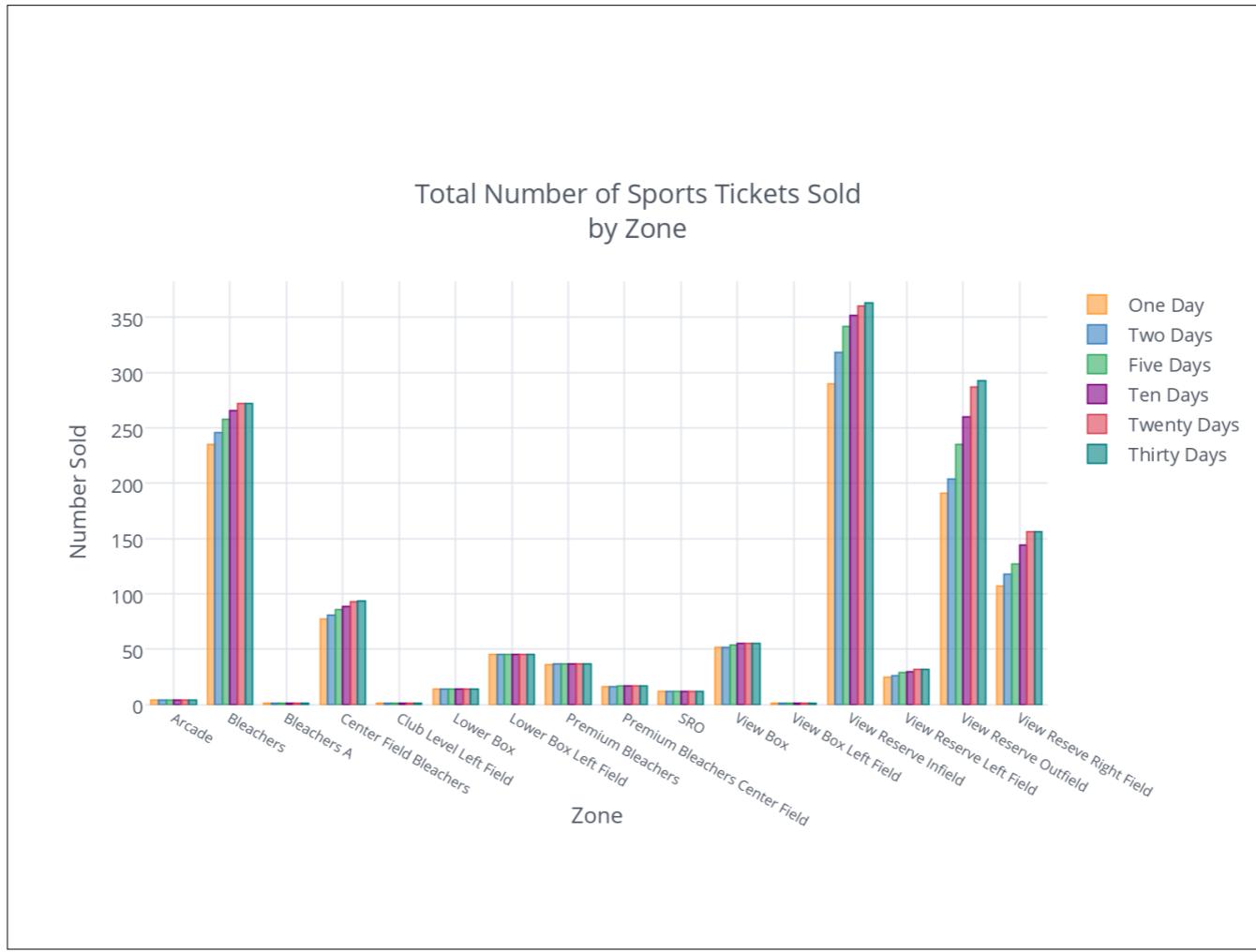
1. Similarly here, Standing Room Only, the cheapest ticket, has managed to turn a slight profit.



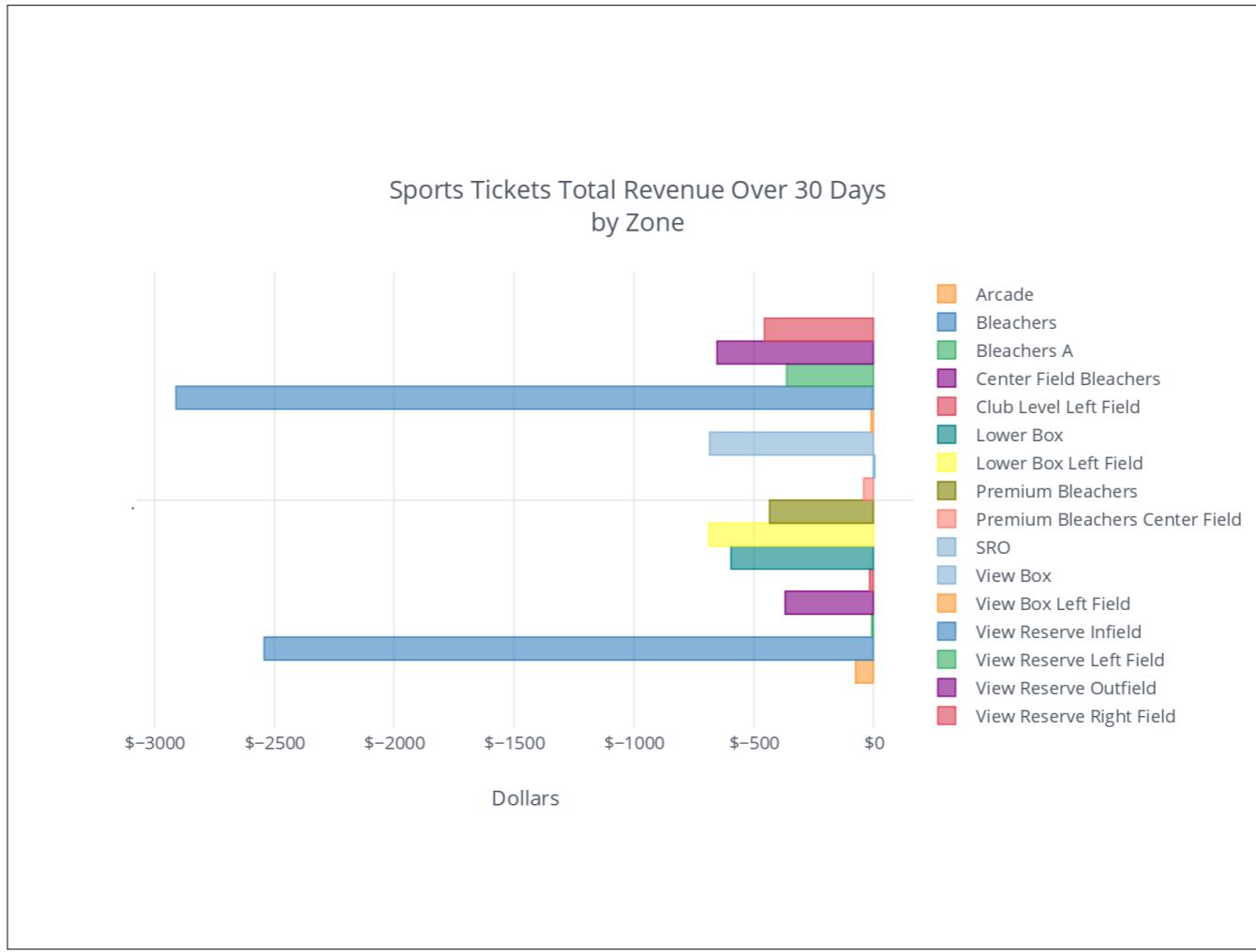
1. For the next few slides, I'll discuss quantity and profit.
2. As you can see, the gross majority of concert ticket sales are in the lower tiers. <next>



1. If we consider the quantity of ticket transactions within just 30 days, the cheapest zone, has proven to show substantial profits.
2. Nothing comes close. <next>



1. Here are the sports ticket sales.
2. Anything prefaced with 'View Reserve' would be commonly referred to as the nose bleed seats.
3. All bleacher seats are at the far end of the baseball diamond. <next>



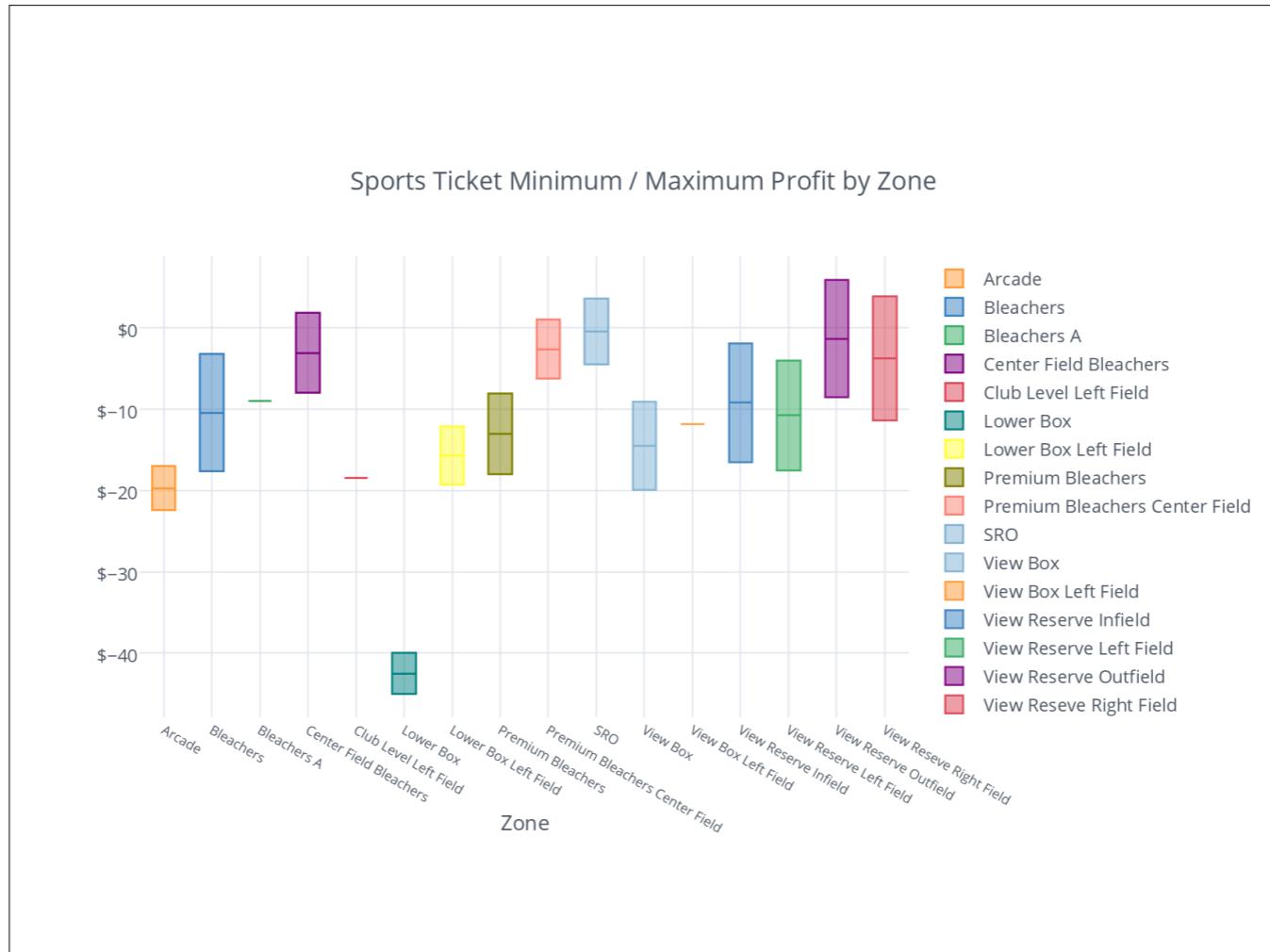
1. But, there is little, if not any profits turned within the sports ticket resales. <next>

# What about zones?

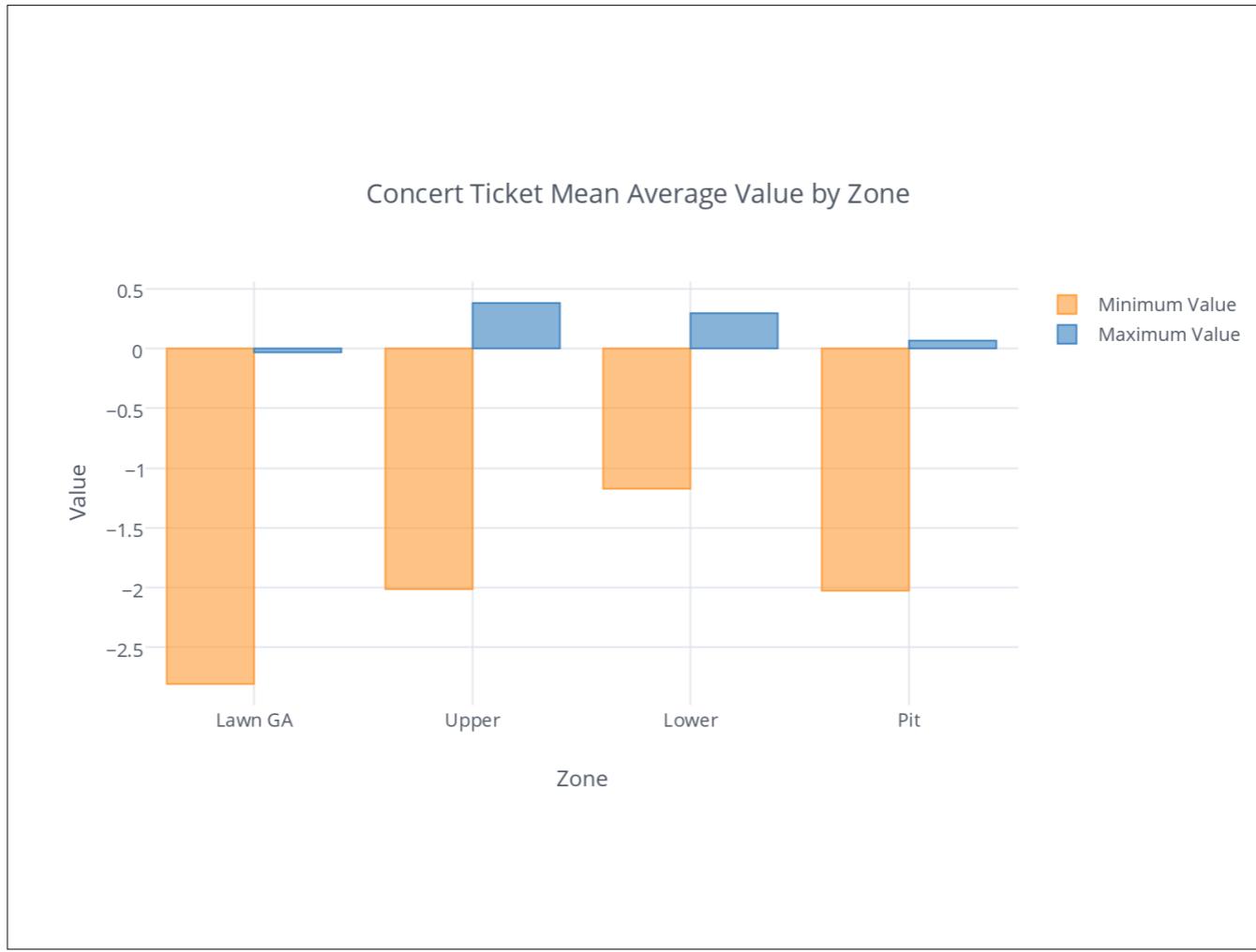
1. But, what about zones?
2. Within each listing are zones for the venue, and within those zones are sections.
3. We're really only going to talk about zones, and I'll explain why in just a moment. <next>



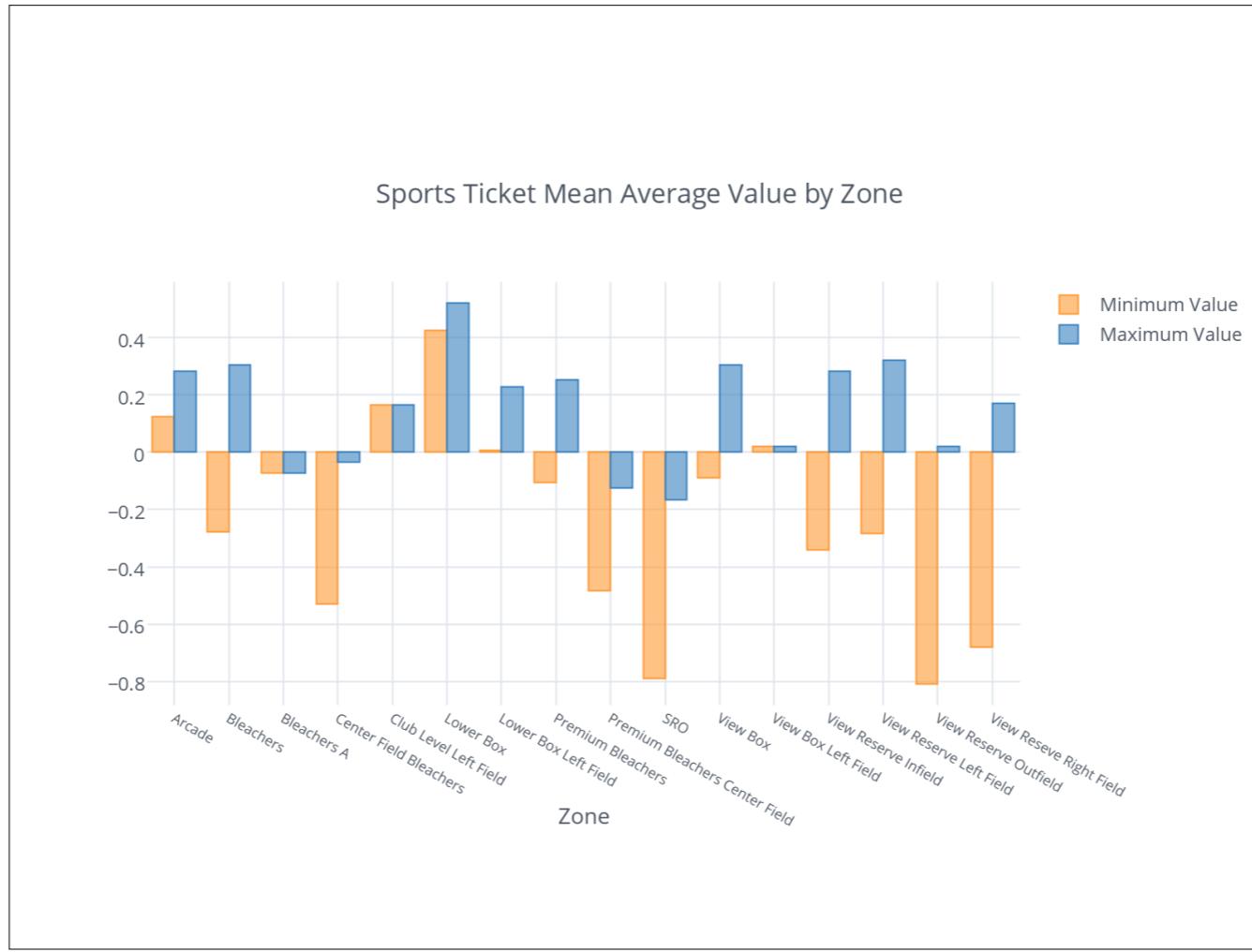
1. For concert tickets, note that the mean average profitability of the most inexpensive zone is close to face value. <next>



1. The sports tickets are erratic and almost all with a mean average below \$0. <next>



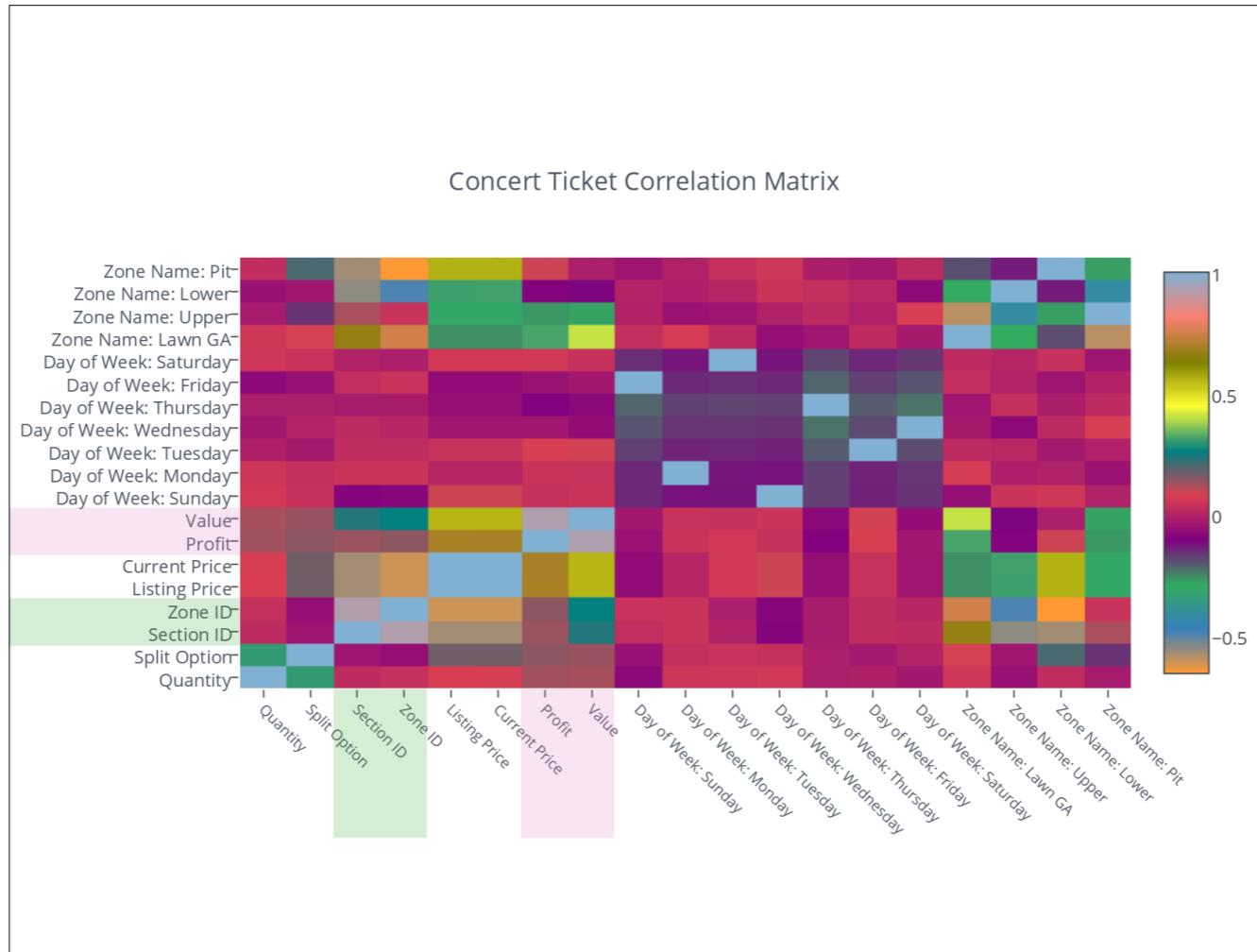
1. Coincidentally, there's a lot of crap listings within the cheapest zone.
2. But, within all of the poor pricing is an opportunity to yield positive returns.
3. Although the Upper and Lower zones afford the most value, they stink when it comes to profitability. <next>



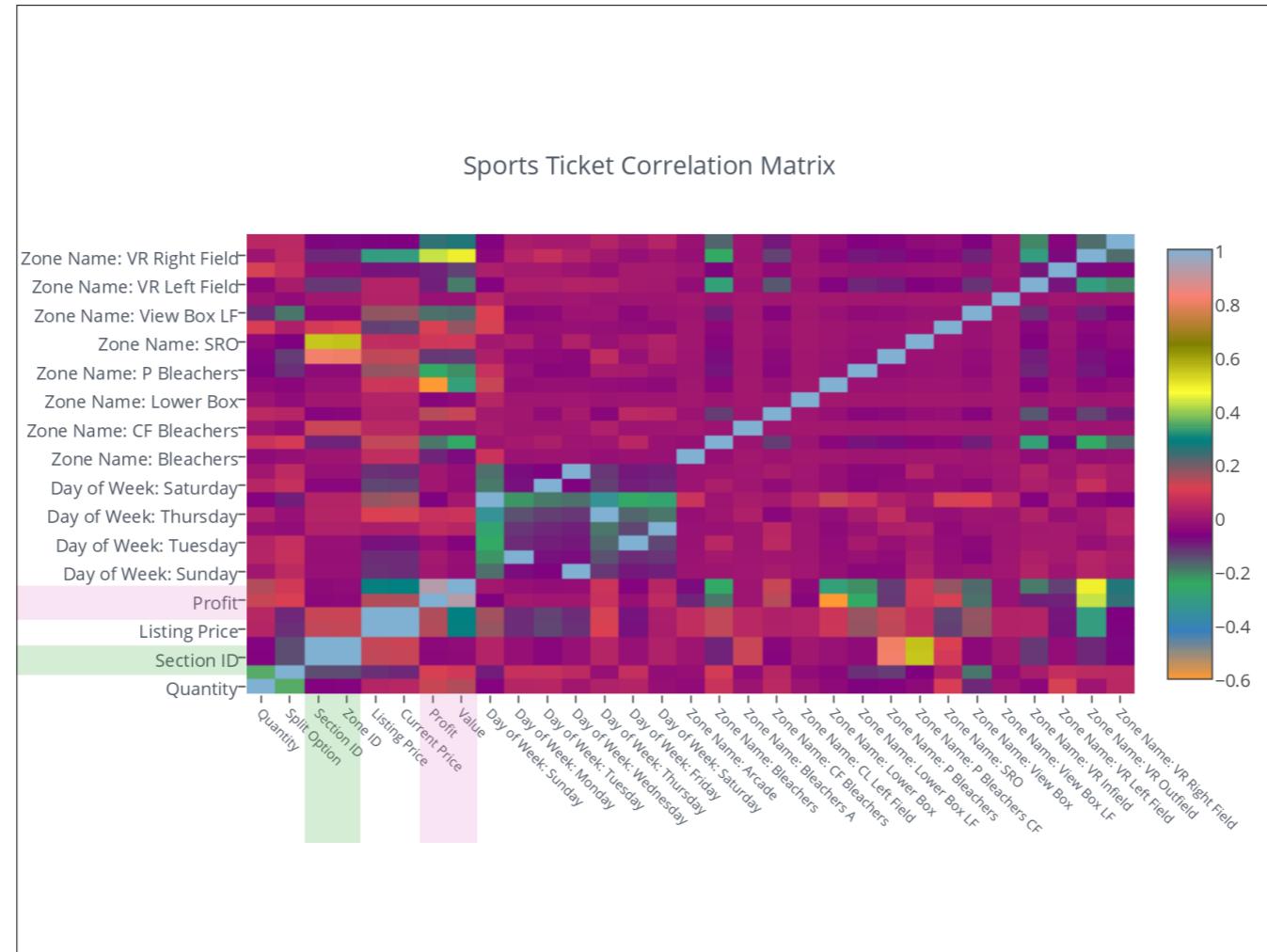
1. The value scale for the sports tickets shows the potential for good deals, but there's unfortunately limited profitability. <next>

# But, what about sections?

1. So, I noticed a problem when I investigated the relationships between the different features of these datasets. <next>



1. Here's a correlation matrix for the dominant features of the concert database.
2. Red to green shows negative to no correlation.
3. Yellow, orange and light blue show a positive correlation.
4. I've highlighted our targets in pink and the zone and section ID features in green.
5. Their predictive capacity toward value and profit is very similar. <next>



1. The same can be said within the sports tickets.
  2. I isolated the zone and section IDs, and...<next>



1. As you can see they are highly collinear, meaning that there is a linear relationship between these two explanatory variables.
2. So, knowing that the zones describe collective amounts of sections within a venue, I removed the sections. <next>

## **2. Personal Goldmine**

**How can I make money off of this?**

1. Phase 2: Personal Goldmine <next>

## Phase 2: Personal Goldmine

- Whoa! I can make money doing this!!?
- Does anyone else know?
- How long can I realistically pull this off?

1. I quickly realized that I could maybe make some money for myself doing this.
2. There is legitimate capital in identifying the most valuable tickets amongst sales trends.
3. Particularly for medium to large scale music events.
4. And, admittedly, I like making money for myself.
5. So, how do I do this? <next>

# Regression Tools

- Linear Regression      
- Adaptive Boosting      
- Random Forest      
- Bayesian Ridge      
- K-Nearest Neighbors      

1. Sorry, I can't make a deck without at least a few emojis.
2. The tools in our toolbox are:
3. Linear Regression: Which takes a bunch of data, and attempts to find the relationship between the independent variable, which is the result, and the dependent variables, the causes.
4. Adaptive Boosting: Hmm... Think of a group of people. Person A, who is the poorest at learning, starts to learn problem X. Whatever part of problem X Person A is not good at, Person B will learn. Whatever Persons A and B are not good at, Person C learns, and so on. Each learner specializes in the weakest area that needs the most improvement. I'm using the default base estimator, a Decision Tree Regressor.
5. Speaking of decision trees:
6. Random Forest: Start with the idea of decision tree. This is a way of making a choice based on a flow chart. The random forest is a way of making a decision based on votes from many decision trees which are created by using subsets of the attributes.
7. Bayesian Ridge: With this we fit a model to optimize the regularization parameters lambda, which is the precision of the weights, and alpha, the precision of the noise.
8. K-NearestNeighbors: A simple algorithm that stores all available cases and classifies new cases based on a similarity measure. We're using it for regression here, which means the distances are averages. <next>

# Regression Tools

- Linear Regression



- Adaptive Boosting



- Random Forest



- Bayesian Ridge



- K-Nearest Neighbors



1.

We'll see scores for all of these, but, in the end, my focus was on Adaptive Boosting. <next>

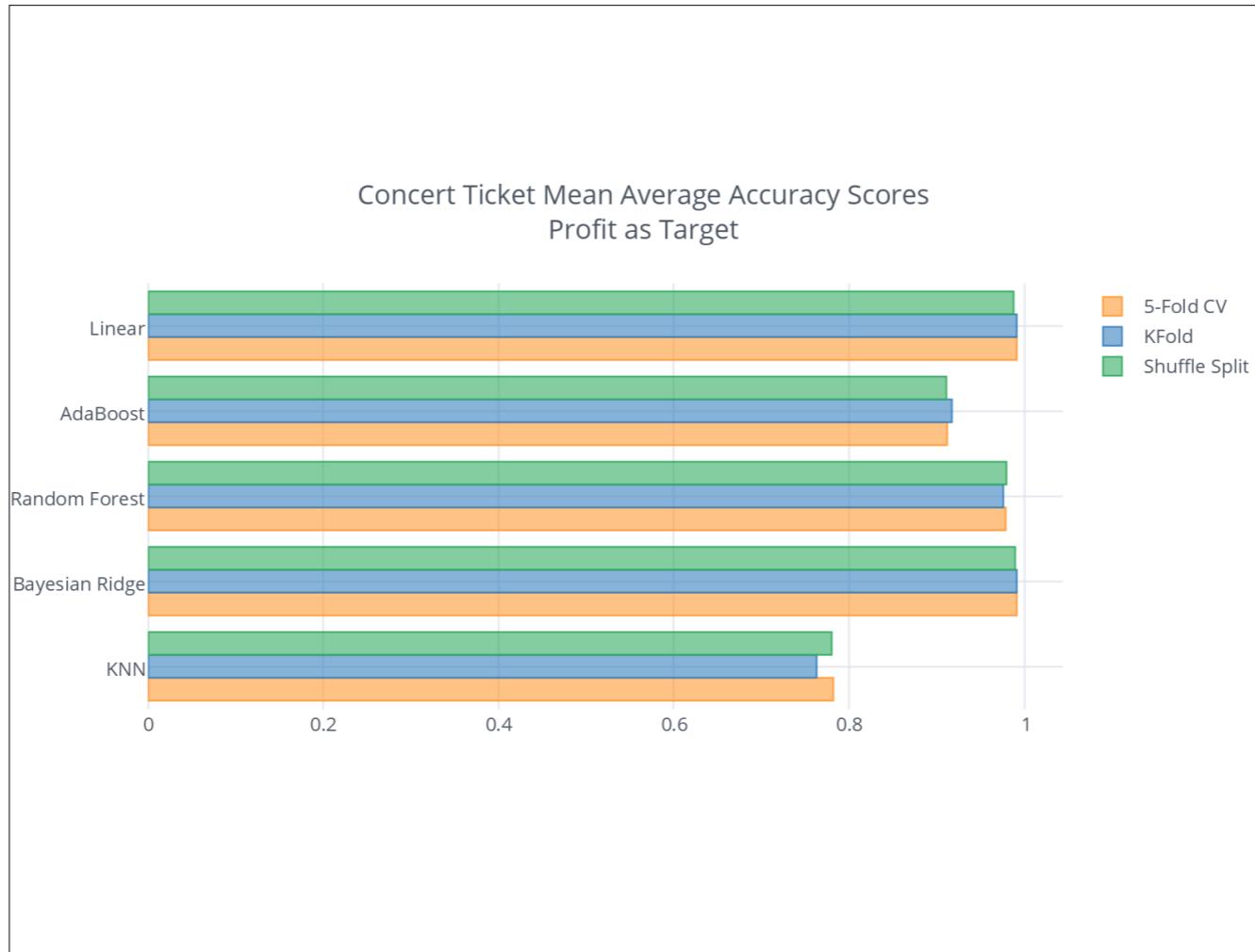
# What are our ingredients?

- Ticket Quantity      Profit:
- Zone ID                Sale Price
- Day of Week           Value:
- Listing Age           Sold Price

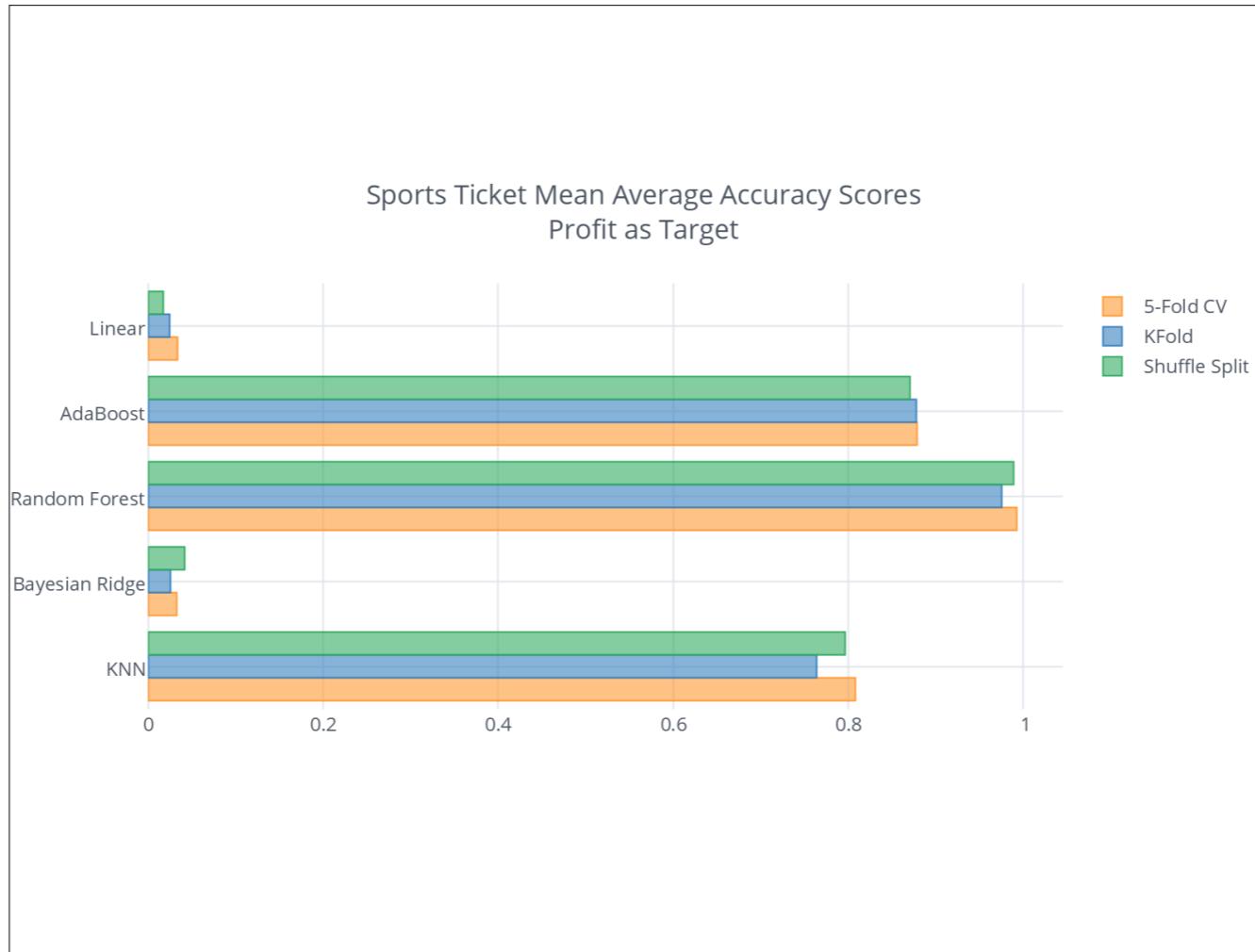
1. So, what are the features of the models?
2. The original dataset had 22 features.
3. This has been pared down to two sets of twelve, with each day of the week as an individual category.
4. Because we have two distinct targets, Sale Price and Sold Price are used respectively.
5. All models were evaluated using a 50/50 train/test split.
6. My goal here was to engineer a standard model that would perform effectively regardless if the target is profit or value. <next>

# First Step: Profit

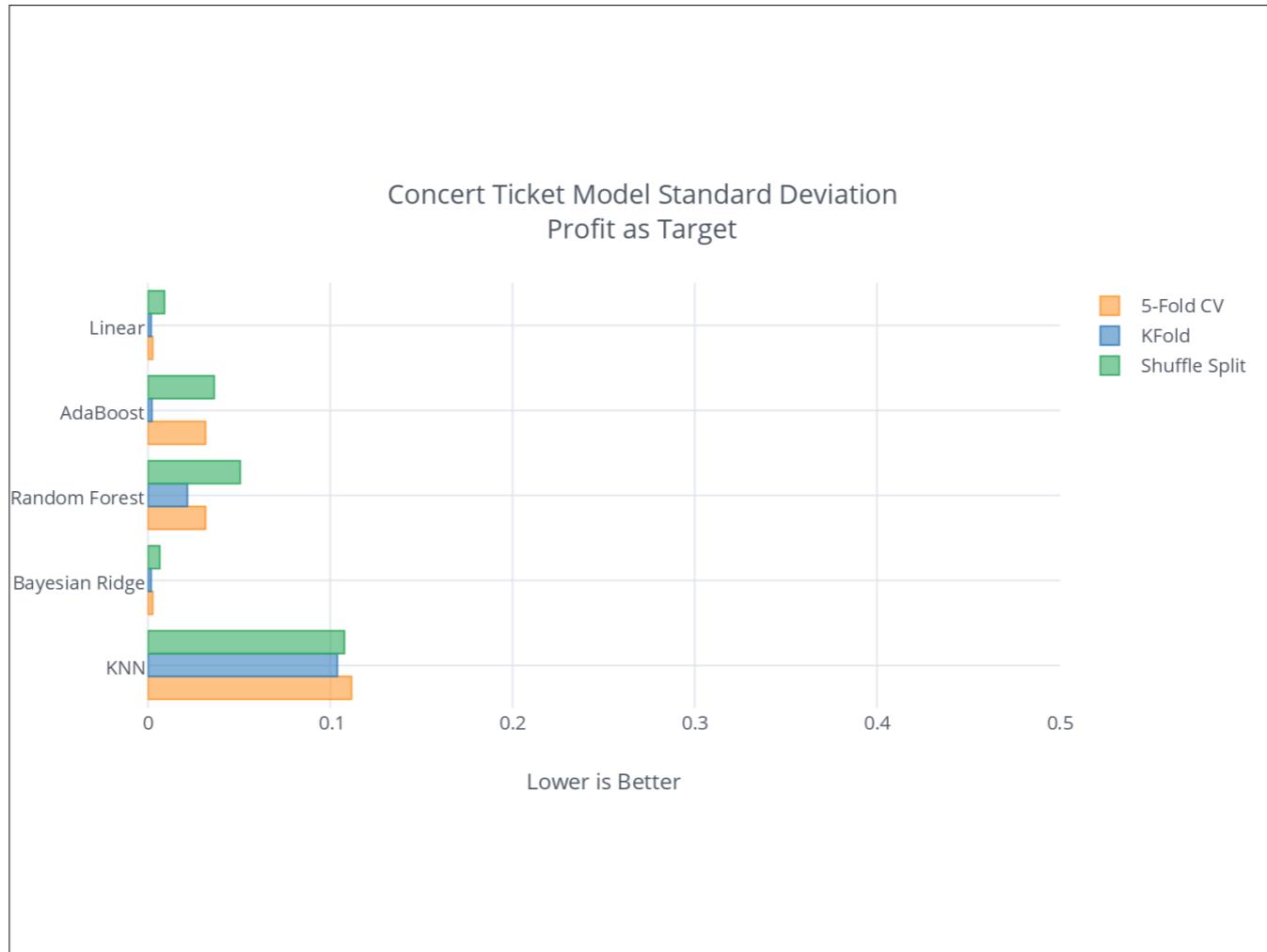
1. First, Profit! <next>



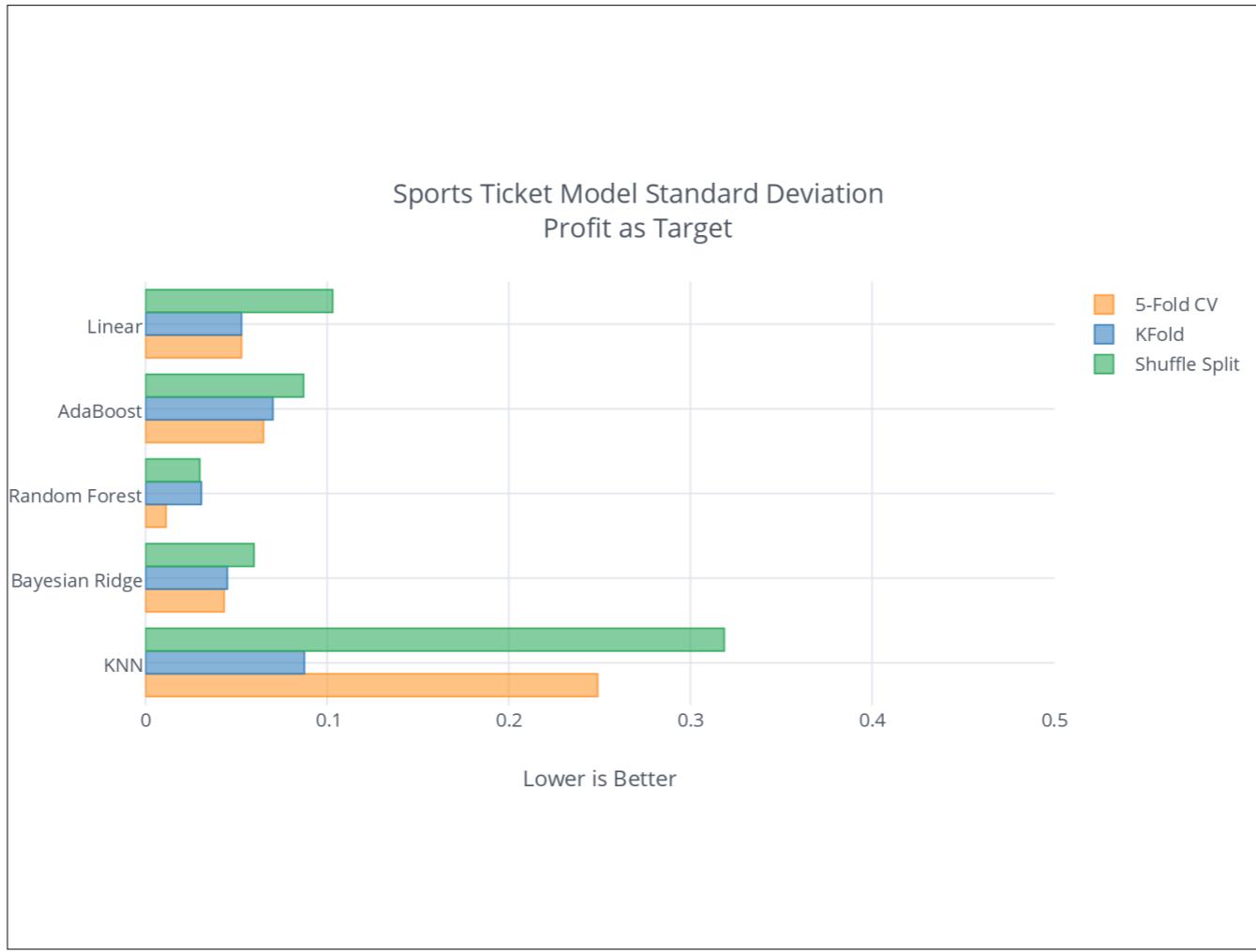
1. Over the next few slides we'll be discussing the accuracy measurements of these different models.
2. I've used three different scoring mechanisms to get a better idea of effectiveness.
3. As you can see, Linear Regression and Bayesian Ridge have overfit the data, which is to say, they are really only successful at predicting this data set.
4. Adaptive Boosting hits the mid-90s range, which I see as the sweet spot. <next>



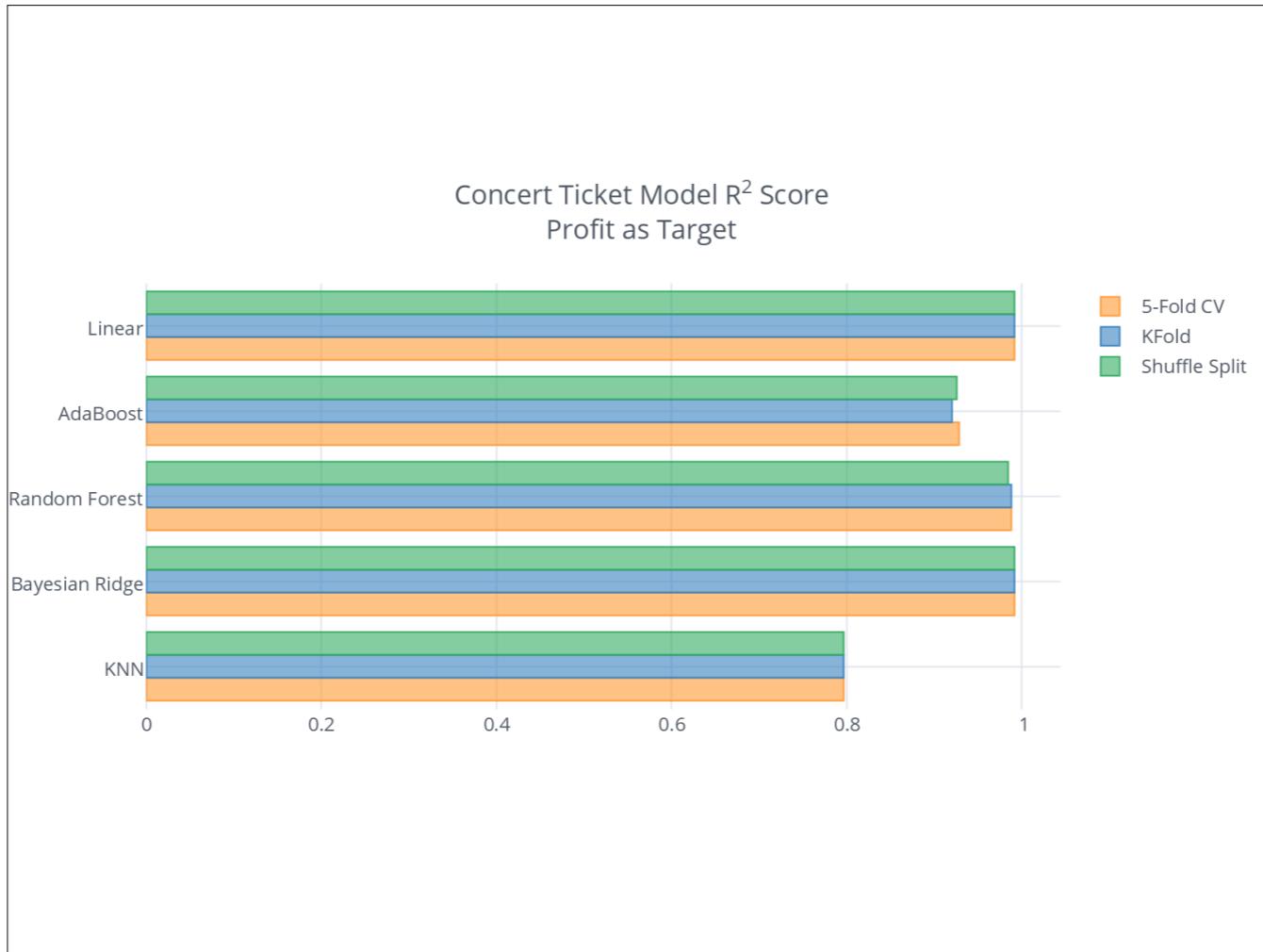
1. And here are the accuracy scores for sports tickets.
2. Linear Regression and Bayesian Ridge did abysmally, but Adaptive Boosting maintains an accuracy in the 90s. <next>



1. Here, we have the standard deviation for our models.
2. A low standard deviation indicates that the data points tend to be close to the expected value of the set. <next>



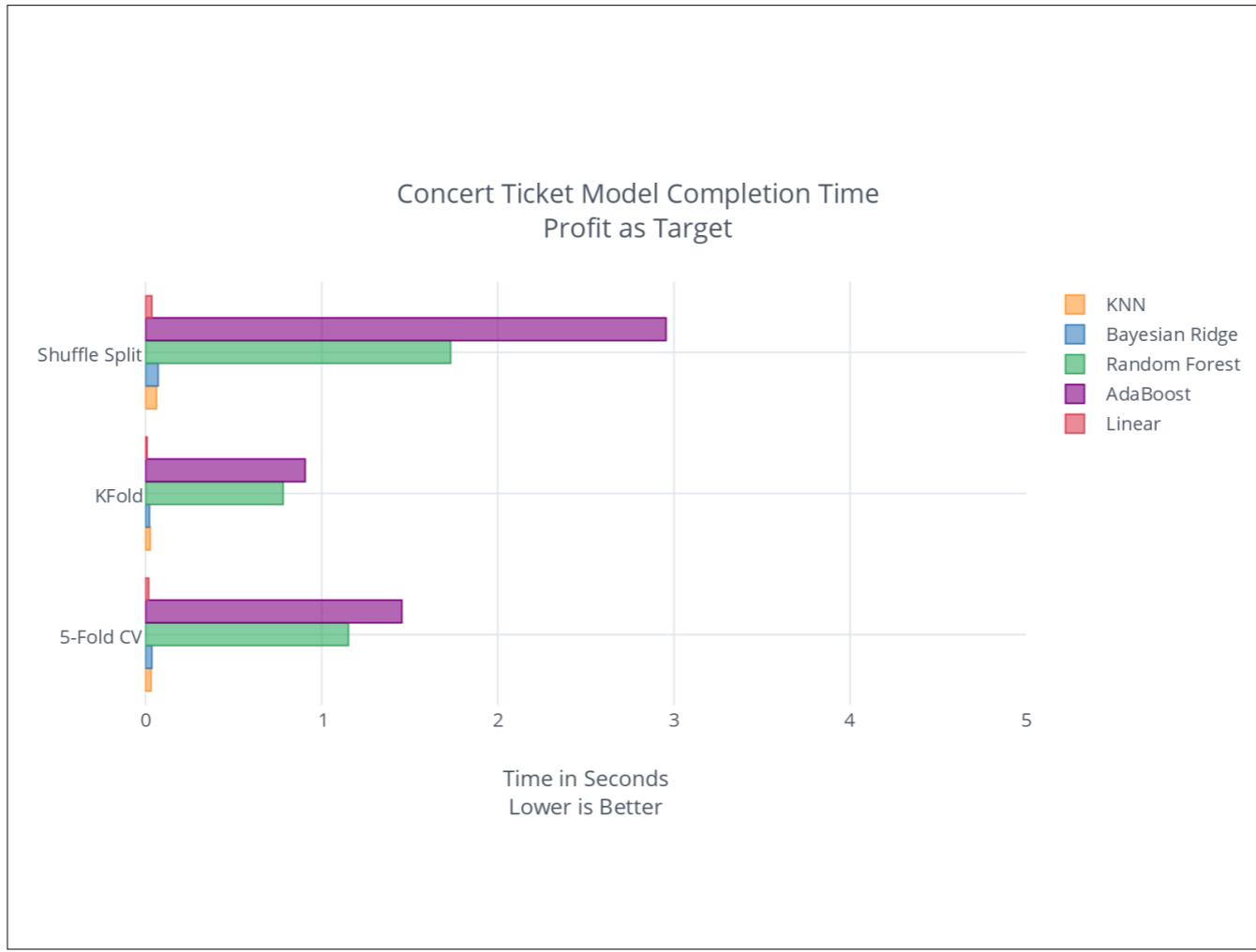
1. And you'll see an increased standard deviation amongst the sports tickets. <next>



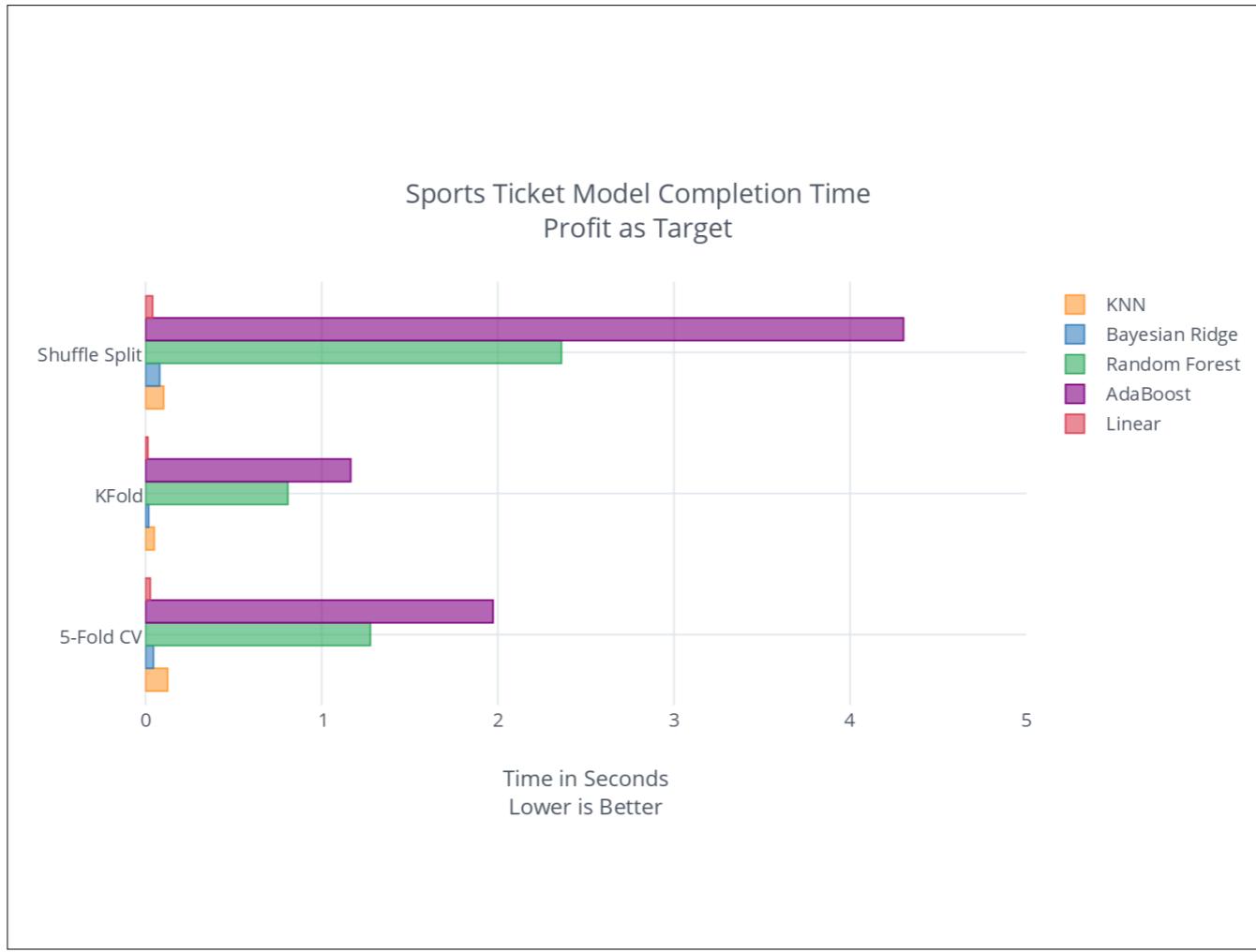
1. R-squared is the “percent of variance explained” by the model.
2. That is, R-squared is the fraction by which the variance of the errors is less than the variance of the dependent variable.
3. It’s called R-squared in a simple regression model, because it is just the square of the correlation between the dependent and independent variables, and that’s commonly denoted by the letter “r”.
4. Here again, I’m most satisfied with Adaptive Boosting’s performance, because it reflects scores in the mid-90s.
5. Let’s check out sports. <next>



1. No surprises here, considering the accuracy of these models.
2. Adaptive Boosting is once again the winner. <next>



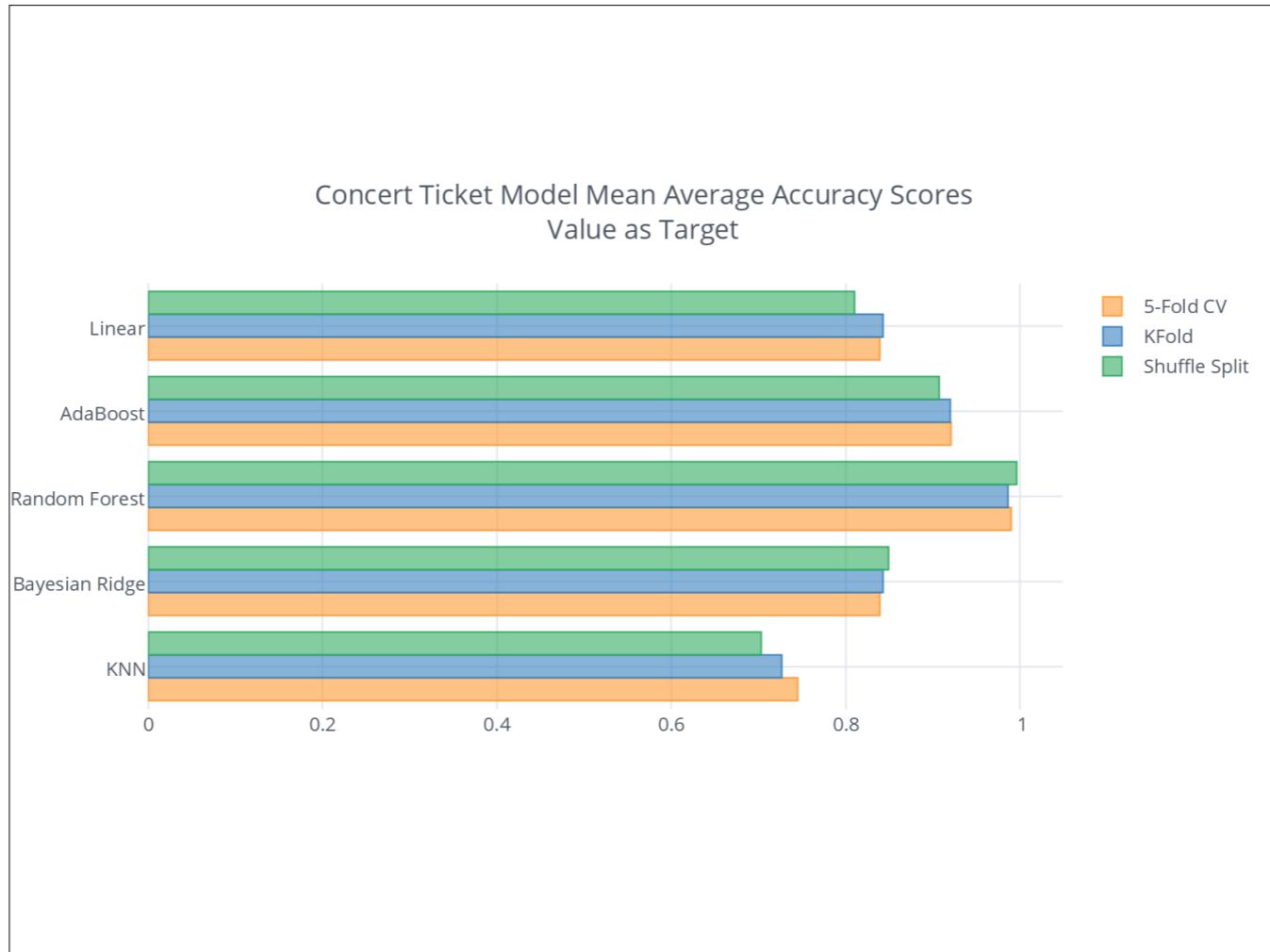
1. For large-scale deployment of these models, the execution time must also be considered.
2. Although Adaptive Boosting worked best at predicting our models, it scores poorly in speed. <next>



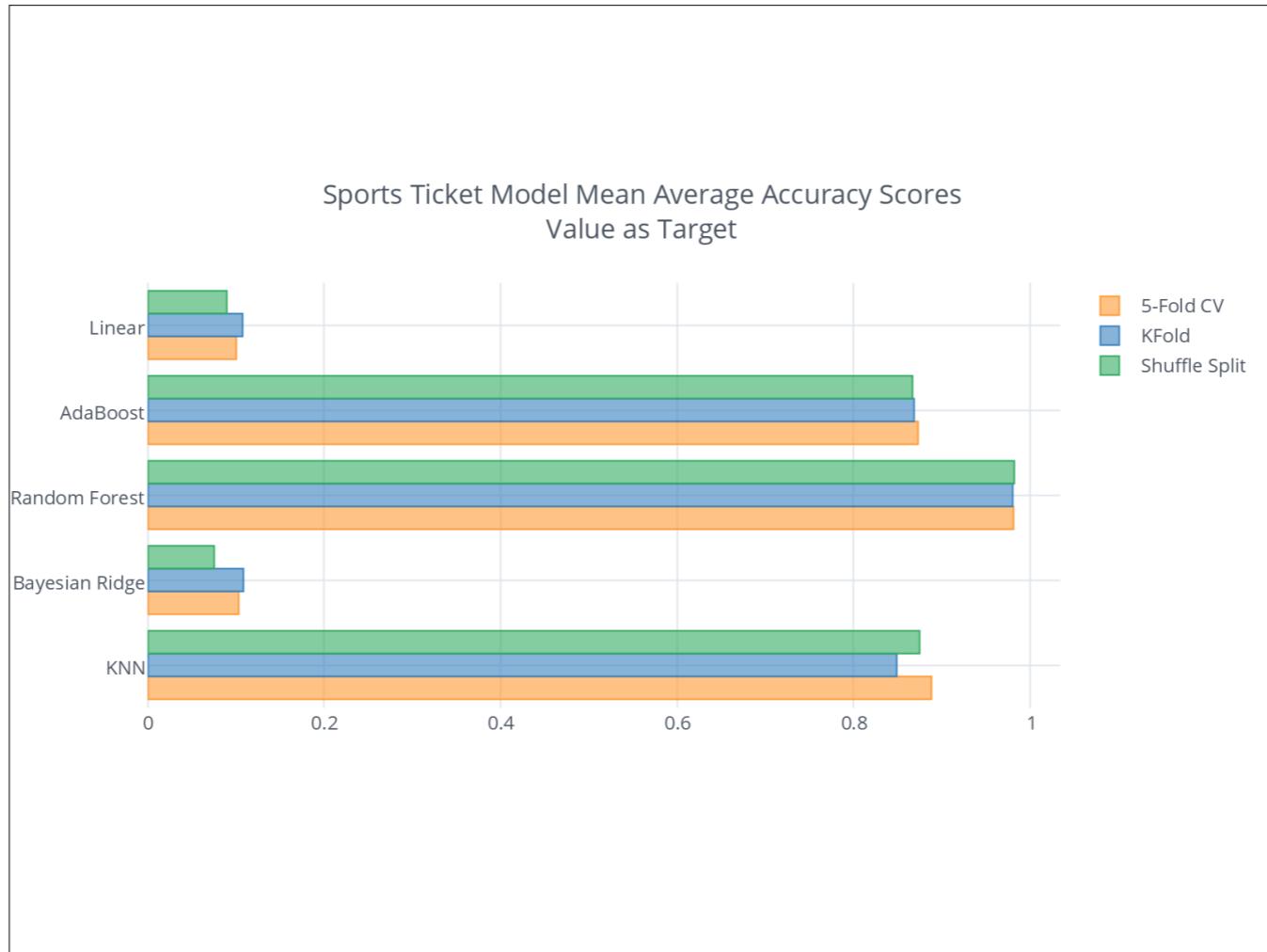
1. Here, too, Adaptive Boosting performs poorly.
2. This is definitely something to keep in mind should this concept be put into production. <next>

## Step 2: Value

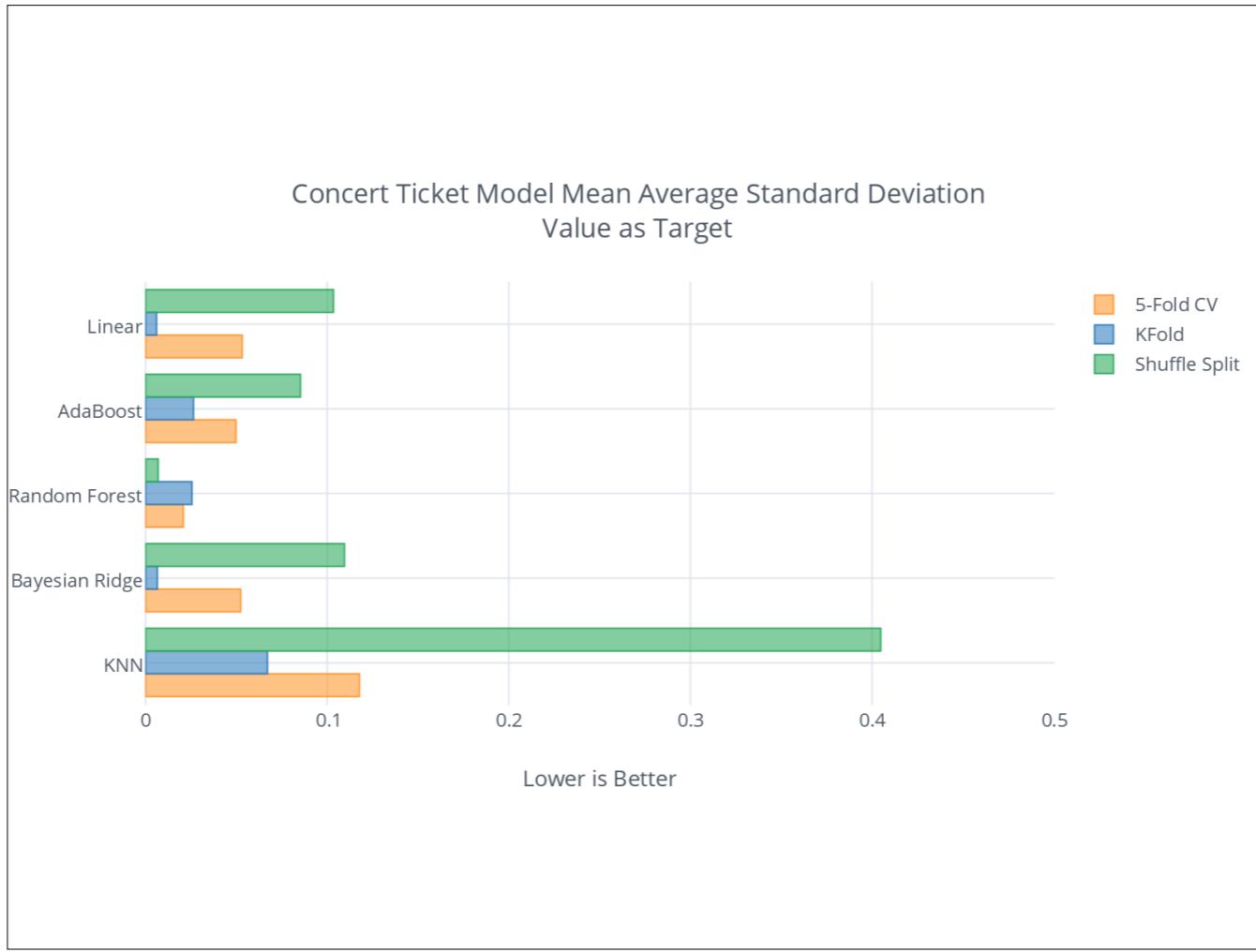
1. Next up, value. <next>



1. The accuracy scores here are more reserved than the concert tickets, with only Random Forest overfitting. <next>



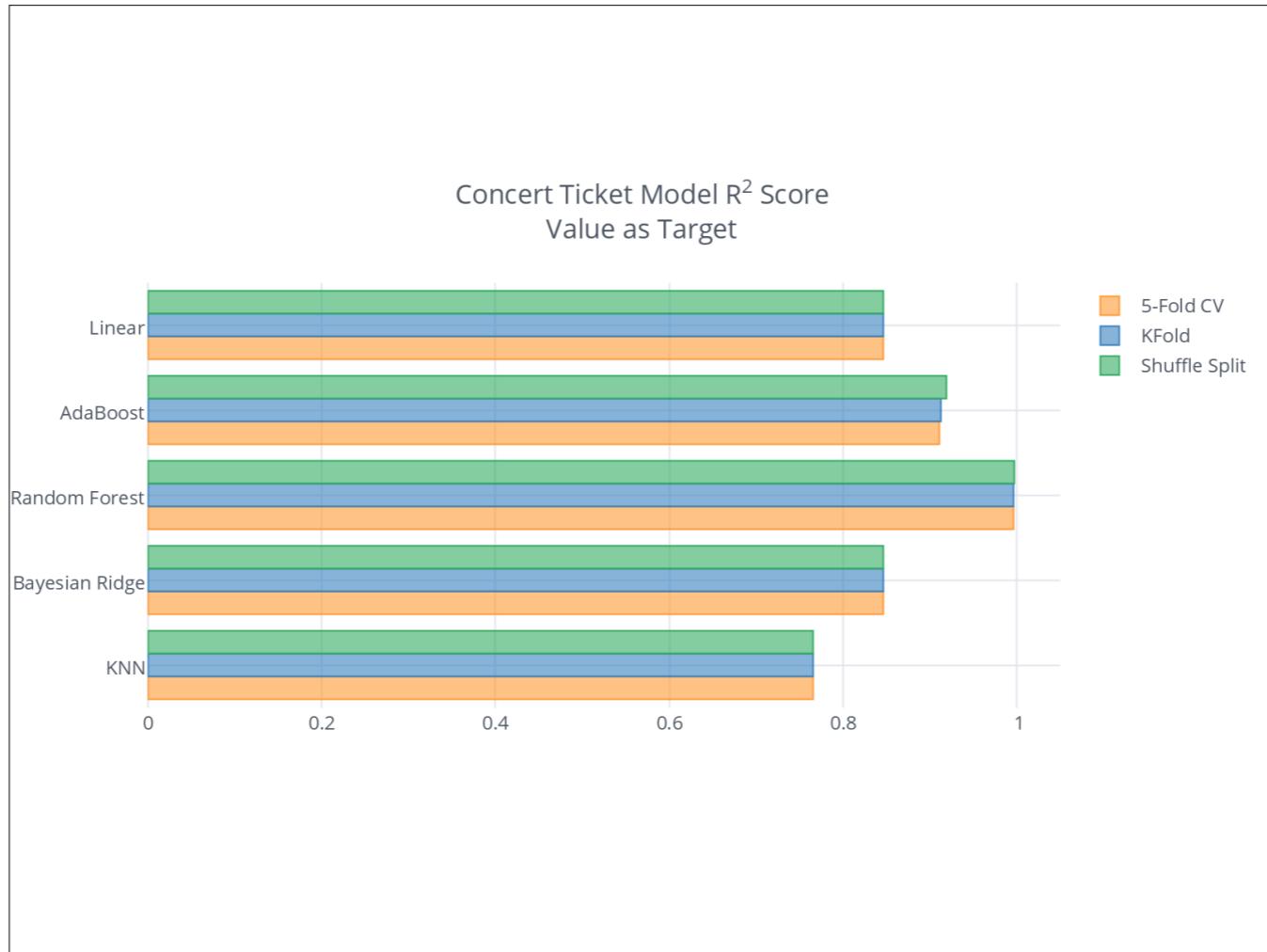
1. But, the value accuracy for sports tickets is very similar to the profit accuracy. <next>



1. Our standard deviation here is similar to evaluating profit as a target, with change primarily in the K-Nearest Neighbors model. <next>



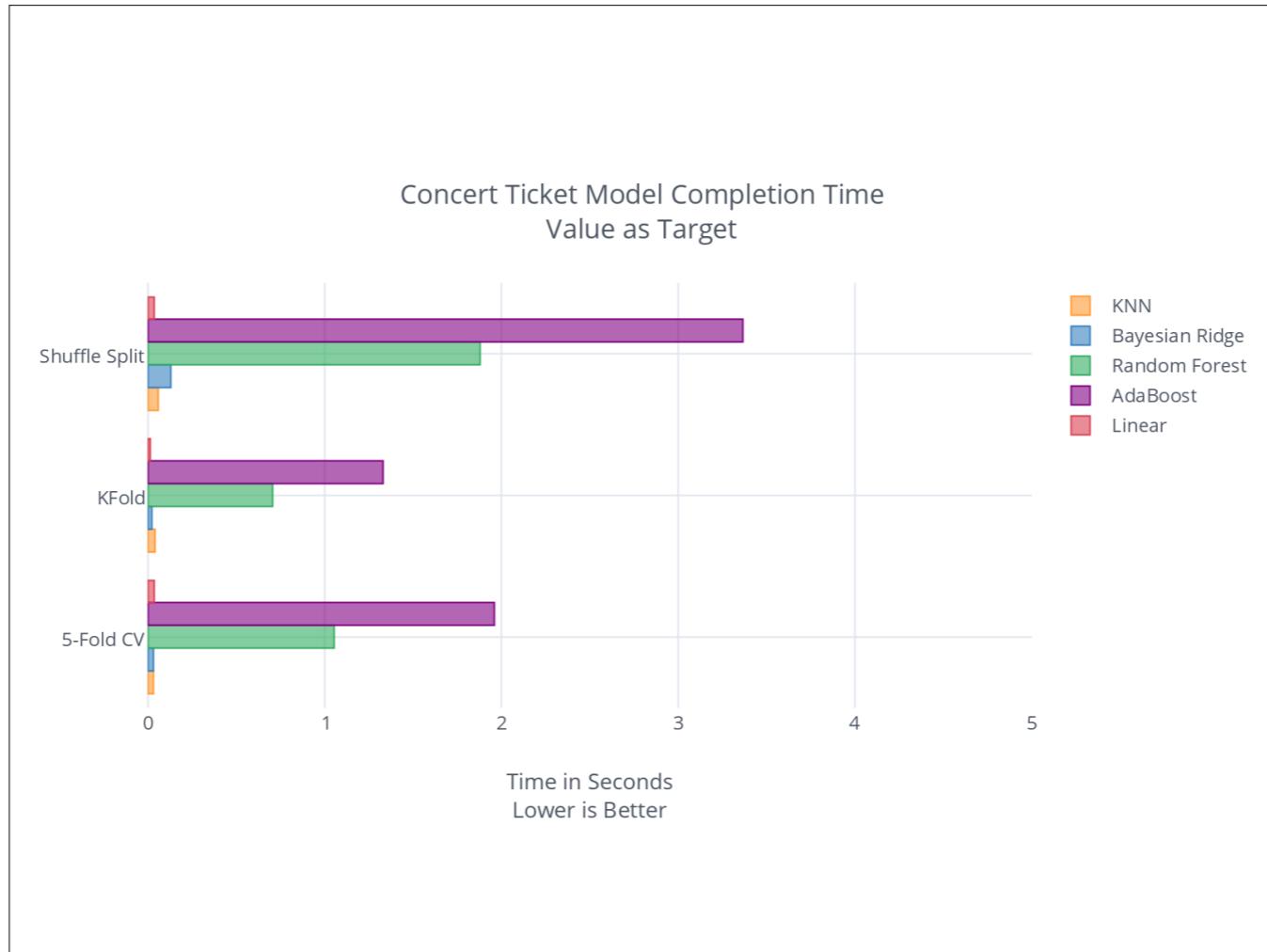
1. But there is a decrease across the board with value as the target. <next>



1. The R-squared scores for concert tickets look like a copy of the value accuracy scores. <next>



1. As do the R-squared scores for the sports ticket models. <next>



1. The execution speed for these models are also highly similar to the scoring for the concert ticket profit models. <next>



1. And the same can be said for the sports ticket value models. <next>

Back to that whole  
‘legality’  
thing.

1. So, let's go back to that whole ‘legality’ thing. <next>

# Legal Schmegal

## **CA Penal Code § 346**

Any person who, without the written permission of the owner or operator of the property on which an entertainment event is to be held or is being held, sells a ticket of admission to the entertainment even, which was obtained for the purpose of resale, at any price which is in excess of the price that is printed or endorsed upon the ticket, while on the grounds of or in the stadium, arena, theater, or other place where an event for which admission tickets are sold is to be held or is being held, is guilty of a misdemeanor.

1. So, is all of this legal?
2. For the most part, yes.
3. There's no federal law barring scalping, save for banning bots, but 15 states have a ban specifically for scalping.
4. California has this: PC 346
5. Here, let me highlight the important part for you. <next>

# Legal Schmegal

## CA Penal Code § 346

Any person who, without the written permission of the owner or operator of the property on which an entertainment event is to be held or is being held, sells a ticket of admission to the entertainment even, which was obtained for the purpose of resale, at any price which is in excess of the price that is printed or endorsed upon the ticket, while on the grounds of or in the stadium, arena, theater, or other place where an event for which admission tickets are sold is to be held or is being held, is guilty of a misdemeanor.

1. So, as long as I'm not anywhere near the event, I should be okay. (crosses fingers)
2. But, there is one other hurdle, and that's StubHub's API license agreement. <next>

## StubHub, Inc. User Agreement

### 2. ABUSING STUBHUB

When using the Site or the Services, you agree that you will not do any of the following:

- Contact or invite contact with other StubHub users for any reason other than the purpose for which you received the StubHub user's contact information or solicit sales outside of StubHub;
- Violate any venue or event promoter rules at events or violate any applicable third party terms of service (for instance, when using our mobile applications);
- Use our Services if you are not able to form legally binding contracts (for example if you are under 18);
- Use the Site or the Services for unlawful purposes or in an unlawful manner;
- Use StubHub's trademarks without our written permission;
- Copy, reproduce, reverse engineer, modify, decompile, disassemble or otherwise attempt to derive source code from, create derivative works from, distribute, or publicly display any content (except for your information) or software from our Site or Services without the prior express written permission of StubHub and the appropriate third party, as applicable;
- Commercialize any StubHub application or any information or software associated with such application;
- Export or re-export any StubHub application or tool except in compliance with the export control laws of any relevant jurisdiction and in accordance with posted rules and restrictions;
- Use any robot, spider, scraper, or other automated means to access our Site for any purpose without our express written permission;
- Take any action that imposes or may impose (to be determined in our sole discretion) an unreasonable or disproportionately large load on our infrastructure or otherwise interferes with the functioning of the Site;
- Bypass our robot exclusion headers, robots.txt rules or any other measures we may use to prevent or restrict access to our Site or Services; or
- Do anything else that StubHub determines, in its sole discretion, misuses the Site or Services or otherwise negatively impacts our marketplace

I've focused on the two significant parts:

1. For one, I should be fine as long as I don't completely automate this process.
2. Unfortunately, this is entirely at the discretion of StubHub, and given the social connotations for the word, "Scalping," StubHub would likely designate this as having a negative impact on their marketplace.
3. Taken collectively, my ability to sustain this concept for an extended period of time is questionable. <next>

## **3. Product**

**How can I sell this to StubHub?**

1. And that brought me to Phase 3: Product.
2. How can I sell this to Stubhub? <next>

# Phase 3: Product

- An alternative sales platform within StubHub
- An alternative to the existing Fan Protect guarantee
- Money, riches, fame, power...

1. So, philanthropic and selfish schemes aside, I focused on an actual product.
2. A product that could be implemented by StubHub as a means of both minimizing seller friction and buyer anxiety while also making money.
3. StubHub has a pre-existing buyer guarantee.
4. But, a guarantee is basically additional work for your customer wrapped in a promise. <next>

# The StubHub FanProtect™ Promise

1. Does not guarantee like tickets on the same night.
2. Buyers are frequently unaware of fraud until they arrive at the event.



## BUY TICKETS WITH CONFIDENCE:

Get valid tickets to any event or your money back.

We go out of our way to find replacement tickets if there is an issue with your order.

A full refund if your event is cancelled and not rescheduled.



## SELL TICKETS SAFELY AND SECURELY:

It's free to list your tickets.

Set your price and adjust it at any time before your tickets sell.

Fast and hassle-free delivery followed by timely payment.

We'll handle communications with buyers.

1. If you read reviews of StubHub online, the primary fraud complaints aren't about the fraud itself.
2. The complaints are focused on the resolution.
3. Current ticket purchase options through StubHub are an instant download from StubHub, emailed tickets from the seller, or mailed tickets from the seller.
4. The only medium with any rapid certifiability is the immediate download. The others are increasingly sketchy.
5. Also, consider that many of these customers are unaware of the fraud until they attempt to redeem the ticket at the event.
6. That must be an awful experience.
7. So, I came up with this. <next>



**The process:**

Interject within the seller listing funnel by rapidly identifying the most desirable tickets and invite direct sales to StubHub.

1. A StubHub Certified sales platform.
2. The goal is to continuously tune these regression models with the wealth of closed source data maintained by StubHub as a means of rapidly identifying the most desirable tickets and inviting direct sales to StubHub. <next>



StubHub can charge sellers a lowered sale fee as a convenience for simplifying and incentivizing the process.

1. In creating a service separate from the customer-managed transaction, StubHub can create a new monetization platform.
2. The draw for this platform is two fold:
3. For the seller, the expediency and the guaranteed sale.
4. It's an immediate return on investment.
5. For the buyer, an assurance that there is no reason to question the authenticity of the purchase.
6. There is no other transaction than the one directly from StubHub. <next>



Similarly, StubHub can incrementally increase the dynamically priced fees to buyers as an assurance that they will absolutely receive their tickets.

1. In addition, StubHub can monetize the trusted sale concept by increasing the dynamically priced fees currently charged to buyers that choose this program.
2. This can offset the reduced fee to sellers, and, in a way, this becomes a self-sustaining product without burdening the bottom line.
3. On a whole, this broadens the entire transaction surface, from seller to buyer.
4. It becomes a new line of revenue, and it has the potential to mitigate complaints against the company. <next>

# Party People: Party Hard

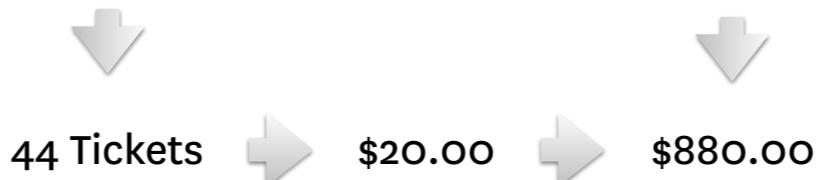
There are over 3,500 festivals and more than 40,000 concerts at over 1,000 venues around the world each year.

Approximately 14,000 of those are medium to large scale events.

1. By LiveXLive.com's count, there are over 3,500 festivals and more than 40,000 concerts at over 1,000 venues around the world each year, with approximately 14,000 of those being medium to large scale events.
2. StubHub currently operates within 47 of those markets worldwide.
3. There's a lot of money to be made, even if we just consider half of those medium to large scale concerts as being targetable by StubHub. <next>

# The Bottom Line

Cheap Seats: 174 \* 25%                      7,000 concerts



\$6,160,000.00

1. Let's use the numbers from this John Legend concert.
2. I'll project that, of the 174 "cheap seats," StubHub could garner a quarter of those as direct resales.
3. I know that's ambitious, but here me out.
4. The average profit from each of those resales is about \$20, and that's without convenience fees for StubHub managing the transaction.
5. At 7,000 concerts, this new line of revenue would equal \$6.16 million dollars.
6. Heck, even if StubHub could only garner 10% of those tickets, that would bring about an additional \$2.44 million dollars. <next>

# Questions

1. This project took a few major turns since the beginning, but it still shows one thing: that there is something here worth continued investigation.
2. And that's what I intend to do.
3. Thank you.