**Step Six: Documentation**

How you express a message is often just as important — if not more important — than the message itself. Even if you find a really valuable insight through data analysis, you’re likely to have a hard time putting that insight into action if you don’t communicate your thoughts about it properly. That’s where storytelling comes in — when you organize your insights into a good story and tell that story effectively, you’ll be much more likely to have a positive impact on the project you're working on.

Please note that this subunit offers a high-level overview of storytelling — you'll take a deeper dive into this topic in a later unit.

Now that you’ve completed your work, it’s important to properly document it. That way, other stakeholders at your company can easily understand your business insights. This is especially important for convincing non-technical stakeholders to buy into the solutions you propose. Dr. Guy Maskall will close out his series on the DSM with this article.



[Write](https://medium.com/m/signin?operation=register&redirect=https%3A%2F%2Fmedium.com%2Fnew-story&source=---two_column_layout_nav-----------------------new_post_sidenav-----------)

[Sign up](https://medium.com/m/signin?operation=register&redirect=https%3A%2F%2Ftowardsdatascience.com%2Fdocumentation-511c1fcb8fe&source=post_page---two_column_layout_nav-----------------------global_nav-----------)

[Sign In](https://medium.com/m/signin?operation=login&redirect=https%3A%2F%2Ftowardsdatascience.com%2Fdocumentation-511c1fcb8fe&source=post_page---two_column_layout_nav-----------------------global_nav-----------)

[[Towards Data Science](https://towardsdatascience.com/?source=post_page-----511c1fcb8fe--------------------------------)](https://towardsdatascience.com/?source=post_page-----511c1fcb8fe--------------------------------)

Published in

[Towards Data Science](https://towardsdatascience.com/?source=post_page-----511c1fcb8fe--------------------------------)

[[](https://guymaskall.medium.com/?source=post_page-----511c1fcb8fe--------------------------------)](https://guymaskall.medium.com/?source=post_page-----511c1fcb8fe--------------------------------)

[Guy Maskall](https://guymaskall.medium.com/?source=post_page-----511c1fcb8fe--------------------------------)

Jul 15, 2020

·

4 min read

·

# Documentation

## Present the right message to the right people in the right way!

A picture containing electronics, circuit

Description automatically generated

Image by [Gerd Altmann](https://pixabay.com/users/geralt-9301/?utm_source=link-attribution&utm_medium=referral&utm_campaign=image&utm_content=3706562) from [Pixabay](https://pixabay.com/?utm_source=link-attribution&utm_medium=referral&utm_campaign=image&utm_content=3706562)

# What’s in a name?

This step could come under a variety of titles, for example, “Storytelling” or “Communication”. In the introductory overview to this series, we likened this step to CRISP-DM’s “Deployment” stage. We don’t go into deploying machine learning models here, but we can phrase this step in the context of “we designed and trained a model and got some results and this is what you need to know about those results and what to do with them”. It isn’t about echoing every minor detail of the entire process verbatim; that’s what your sequence of Jupyter notebooks is for! It is those that act as your audit trail of the evolution of the process in a reproducible manner. No, by this point, if you were working as a car design engineer, your new car has been designed, engineered, manufactured, and is about to hit the showrooms. You’re writing the sales brochure now. “This baby will do 0–60 in 3 seconds, 50 miles to the gallon, and has lane-change assist technology. This is what lane-change assist technology does, and this is why you want it…” The research backing up the car’s performance stats, and how you developed your new lane-change assist technology is what your previous data science artifacts are for. Oh, and the technical details of the development and implementation of the technology may very likely be a proprietary company secret. Shhhh.

Your data science work will count for little if no one sees the point in using it, if they don’t understand how to use it, or how it can benefit them. Your car was designed to accommodate a family of four. Show how your car makes that holiday journey enjoyable and safe for the children. Show how the boot (alright, y’all — trunk!) securely takes the weekly grocery shopping and keeps it upright despite that sudden braking manoeuvre. Show how its features solve problems for the customer! That’s what the client wants to know, and that’s what will have them adopt your solution and come back to you for more.

# Formats

## The slide deck

There are a variety of formats you may be asked to use for this. First, there’s the slide deck. Slides should not be overly verbose. It’s not a game to try to squeeze as many words as possible onto a slide. You should also avoid equations. There is a balance here, though. When you’re giving a presentation based off your slides, you don’t want them too wordy or else the audience will be trying to read your written paragraphs on the screen and not paying attention to you. On the other hand, it’s quite common for busy execs at C-level to ask for your slide deck in advance for a pre-read prior to your meeting. In that case, the slides definitely need to be understandable by the intended audience when read unaccompanied, rather than being a series of one word bullet points. Remember, it’s not about getting every subtlety, every justification or side issue on the page. What do you really want them to know? When considering a new car purchase, the Joneses are unlikely to care two hoots how clever you were tuning that parameter in the electronic control unit; they will care about a comfortable ride or improved fuel economy. Save your technical pride for your team meetings!

## The report

You may also be asked to write up a report. This is your chance to tell a fuller story in a way that definitely should stand alone (not require you to stand over the reader’s shoulder explaining things). Again, it is not really the place to showcase your equations or deep technical knowledge. The likely audience is the customer, and/or business executive who has a little more time on their hands. The most important advice here, in line with that above, is consider your audience. And also consider the expected length of the report. Those, along with the message(s) you want to make sure you get across, will guide you towards the appropriate technical depth. Everything else is just writing style. And writing style is important! A classic guide is [Strunk and White](https://en.wikipedia.org/wiki/The_Elements_of_Style). It’s not without its critics, but it’s definitely not a bad place to start. You can find others. Your company may very well have its own style guide! You can, and will, develop your own voice.

# About this article

This is the seventh, and final, article of a linked series written to provide a straightforward introduction to getting started with the data science process. You can find the introduction [here](https://medium.com/@guymaskall/the-data-science-method-dsm-35200eb4984), and the previous article [here](https://medium.com/@guymaskall/modelling-e096495d8d70).

“we designed and trained a model and got some results and this is what you need to know about those results and what to do with them”.

Guided Capstone Project Report

Summary of the Recommendations for Big Mountain Resort

This project is based on Big Mountain wanting to increase their current ticket price of $81.

Based on Big Mountain’s distribution in the US market and from the data tested and gathered, the model data has put into consideration the additional operating cost of the chair lift installed  and the basis of each visitor on average buying 5 day tickets. The ticket price for the resort has the potential to increase above $81. The model data suggests a ticket price of $95.87 versus the actual price of $81. Even with the expected mean absolute error of $10.39, there is room for increasing the ticket price that could be supported in the marketplace. Big Mountain has a lot of features/facilities to offer that would support the increase in ticket price.

Let us start by reviewing the first 2 options listed by the business leaders. The first option was to close up to 10 of the least used runs. The second option was to increase the vertical drop by adding a run to a point 150 ft lower down with the installation of an additional chair lift to bring skiers back without additional snow making coverage.

The recommendation is the combination of the two options. The first one, by closing at least one of the runs which will not affect the ticket price or revenue. Losing one run may have operational cost savings that can be applied to the bottom line.

The second option will support an increase in the ticket price by $1.99 and an increase in revenue of $3474638 over a season. There might be a possibility of using one of the potential run and chair lift that the resort already have.

These options are recommendable because it supports the entire project’s motivation of increasing its ticket price which is supported by the data analysis made.

It would also be recommended the need for further investigation into the cost of adding vertical drop, possible redirection of a run and possible new lift and finding out if there is a run that is extendable with property and geographic limitations within the overall property. Aside from these, the need to investigate the operating cost of the resort might be a key data worth looking into.

If the tested model is useful, the business could apply the increase in price as recommended by the data. Any new parameters that are related to the present business problem may be tested but if parameters that are not directly related may be reviewed if it makes sense to add it or make a new project to investigate further. It would be better to make recommendations based on parameters that the business leaders may have used in the past, to base any future business decisions.

Problem Statement

Big Mountain Resort wants to increase their revenue by:

1. Charging a premium above the average price for the use of the resort’s facilities and
2. Cutting the cost without undermining the ticket price

The business problem is and to have data to solve the problem. The business problem was a general one of modeling resort revenue.

The data science problem identified is to predict the adult weekend ticket price for ski resorts.

Recommendation and Key Findings

* Big Mountain has a lot of features/facilities to offer that would support the increase in ticket price.
* the model data has put into consideration the additional operating cost of the chair lift installed
* each visitor on average buying 5 day tickets.
* This project is based on Big Mountain wanting to increase their current ticket price of $81.
* Based on Big Mountain’s distribution in the US market and from the data tested and gathered, the model data has put into consideration the additional operating cost of the chair lift installed  and the basis of each visitor on average buying 5 day tickets. The ticket price for the resort has the potential to increase above $81. The model data suggests a ticket price of $95.87 versus the actual price of $81. Even with the expected mean absolute error of $10.39, there is room for increasing the ticket price that could be supported in the marketplace. Big Mountain has a lot of features/facilities to offer that would support the increase in ticket price.
* Let us start by reviewing the first 2 options listed by the business leaders. The first option was to close up to 10 of the least used runs. The second option was to increase the vertical drop by adding a run to a point 150 ft lower down with the installation of an additional chair lift to bring skiers back without additional snow making coverage.
* The recommendation is the combination of the two options. The first one, by closing at least one of the runs which will not affect the ticket price or revenue. Losing one run may have operational cost savings that can be applied to the bottom line.
* The second option will support an increase in the ticket price by $1.99 and an increase in revenue of $3474638 over a season. There might be a possibility of using one of the  existing runs and chair lifts that the resort already has.
* These options are recommended because it supports the project’s motivation of increasing its ticket price, which is supported by the data analysis made.

Option 1: Effect of Closing Runs on Price and Revenue

The first option was to close up to 10 of the least used runs.

* The model shows closing one run will not affect ticket prices
* Some operational costs can be gained
* the first 2 options listed by the business leaders. The first option was to close up to 10 of the least used runs. The second option was to increase the vertical drop by adding a run to a point 150 ft lower down with the installation of an additional chair lift to bring skiers back without additional snow making coverage.
* The recommendation is the combination of the two options. The first one, by closing one of the runs which will not affect the ticket price or revenue. Losing one run may have operational cost savings that can be applied to the bottom line.
* The second option will support an increase in the ticket price by $1.99 and an increase in revenue of $3474638 over a season. There might be a possibility of using one of the existing runs and chair lifts that the resort already has.
* These options are recommended because they support the project’s motivation of increasing the ticket price, which is supported by the data analysis made.
* It would also be recommended the need for further investigation into the cost of adding vertical drop, possible redirection of a run and possible new lift and finding out if there is a run that is extendable with the property and geographic limitations within the overall property. Aside from these, the need to investigate the operating cost of the resort might be a key data worth looking into.
* If the tested model is useful, the business could apply the increase in price as recommended by the data. Any new parameters that are related to the present business problem may be tested but if parameters that are not directly related may be reviewed if it makes sense to add it or make a new project to investigate further. It would be better to make recommendations based on parameters that the business leaders may have used in the past, to base any future business decisions.

Option 2

The second option was to increase the vertical drop by adding a run to a point 150 ft lower down with the installation of an additional chair lift to bring skiers back.

* supports an increase in the ticket price by $1.99 and an increase in revenue of $3,474,638 over a season.
* ers back without additional snow making coverage.
* The recommendation is the combination of the two options. The first one, by closing at least one of the runs which will not affect the ticket price or revenue. Losing one run may have operational cost savings that can be applied to the bottom line.
* The second option will support an increase in the ticket price by $1.99 and an increase in revenue of $3474638 over a season. There might be a possibility of using one of the potential run and chair lift that the resort already have.
* These options are recommendable because it supports the entire project’s motivation of increasing its ticket price which is supported by the data analysis made.
* It would also be recommended the need for further investigation into the cost of adding vertical drop, possible redirection of a run and possible new lift and finding out if there is a run that is extendable with property and geographic limitations within the overall property. Aside from these, the need to investigate the operating cost of the resort might be a key data worth looking into.
* If the tested model is useful, the business could apply the increase in price as recommended by the data. Any new parameters that are related to the present business problem may be tested but if parameters that are not directly related may be reviewed if it makes sense to add it or make a new project to investigate further. It would be better to make recommendations based on parameters that the business leaders may have used in the past, to base any future business decisions.

Modeling results and Analysis

Big Mountain Resort modelled price is $95.87, actual price is $81.00.

Even with the expected mean absolute error of $10.39, this suggests there is room for an increase.

The validity of our model lies in the assumption that other resorts accurately set their prices according to what the market supports.

If the model is useful to the business, then they could apply the increase in price as recommended by the data. Any new parameters that are related to the present business problem may be tested but if parameters that are not directly related may be reviewed if it makes sense to add it or make a new project to investigate further.

It would be better to make recommendations based on parameters that the business leaders may have used in the past to base any future business decisions.

The random forest model has a lower cross-validation mean absolute error by almost $1. It also exhibits less variability. Verifying performance on the test set produces performance consistent with the cross-validation results.

Now that we have a model for ski resort ticket price and leverage it into what Big Mountain facilities might actually support and explore the sensitivity of changes to various resort parameters. All this with the assumption that prices are set by a free market and how much people values certain facilities.

Set about calculating the price based only on Big Mountain’s competitors.

Big Mountain Resort modelled price is $95.87, actual price is $81.00.

Even with the expected mean absolute error of $10.39, this suggests there is room for an increase

This result should be looked at optimistically and doubtfully! The validity of our model lies in the assumption that other resorts accurately set their prices according to what the market (the ticket-buying public) supports. The fact that our resort seems to be charging that much less that what's predicted suggests our resort might be undercharging. But if ours is mispricing itself, are others? It's reasonable to expect that some resorts will be "overpriced" and some "underpriced." Or if resorts are pretty good at pricing strategies, it could be that our model is simply lacking some key data? Certainly we know nothing about operating costs, for example, and they would surely help.

If this model is useful to the business, then they could apply the increase in price as recommended by the data. Any new parameters that are related to the present business problem may be tested but if parameters that are not directly related may be reviewed if it makes sense to add it or make a new project to investigate further.

It would be better to make recommendations based on parameters that the business leaders may have used in the past, to base any future business decisions.

Summary and Conclusion

* price of $81.
  + there is room to increase the price based on the mean absolute error of $10.39.
  + The ticket price can be supported by the market with the facilities that the resort offers.
* Reduce operational costs by closing one run
* Increase the ticket price (+$1.99) and revenue for the resort by increasing the vertical drop to 150 ft.
* Room to continue investigating other related parameters, such as operating costs, to increase the resort’s revenue.

This project is based on Big Mountain wanting to adjust its pricing. The problem is addressed by training a model to predict Big Mountain's ticket price based on data from all the other resorts. We don't want Big Mountain's current price to bias this, a price was calculated based only on its competitors.

calculating expected Big Mountain ticket price from the model and this was performed using the model predict method. Big Mountain Resort modelled price is $95.87, actual price is $81.00. Even with the expected mean absolute error of $10.39, this suggests there is room for an increase.

I would recommend a combination of scenario 1 and 2. Scenario 1 tells us that by closing one run will not affect the ticket price or revenue. But losing one run may also have operational cost savings that can be applied to the bottom line. Modifying a run when considering scenario 2 to adding a vertical drop of 150 ft. This indicates support for an increase in ticket prices and revenues. This scenario will also entail additional investigation into the cost of adding more vertical drop, possible redirection of a run and answer the question of the feasibility of physically adding the drop needs to be investigated.

Summary/Conclusion

This project is based on Big Mountain wanting to adjust its pricing. The problem is addressed by training a model to predict Big Mountain’s ticket price based on the data from all other resorts. We ended up with a modeled price of $95.87 versus actual price of $81. It means that there is room to increase the price based on the mean absolute error of $10.39. A ticket price that can still be supported by the market with the facilities that the resort offers.

With this in mind, the combination of the two mentioned options seems to have the potential to have operational cost savings  by closing one run and increase the ticket price and revenue for the resort by increasing the vertical drop to 150 ft. There is also room to continue investigating other related parameters like operating cost of the resort in order to increase the resort’s revenue in the long run.

1

**Recap**

You've made it through the entirety of the DSM — well done! This unit was all about learning this structured approach to solving data science problems. You're now set up to be able to tackle the DSM on your own, which you'll do while working on your second capstone. The next unit takes a much closer look at the second step of the DSM: data wrangling. You'll also come up with ideas and a project proposal for your second capstone. Let's get to it!