

The Fine Line Between Predictive and Prescriptive Analytics (with examples)

And how to climb the stairs to the state-of-art of Analytics



Victor Almeida · [Follow](#)

Published in Geek Culture

8 min read · Apr 14, 2021



Listen



Share

... More

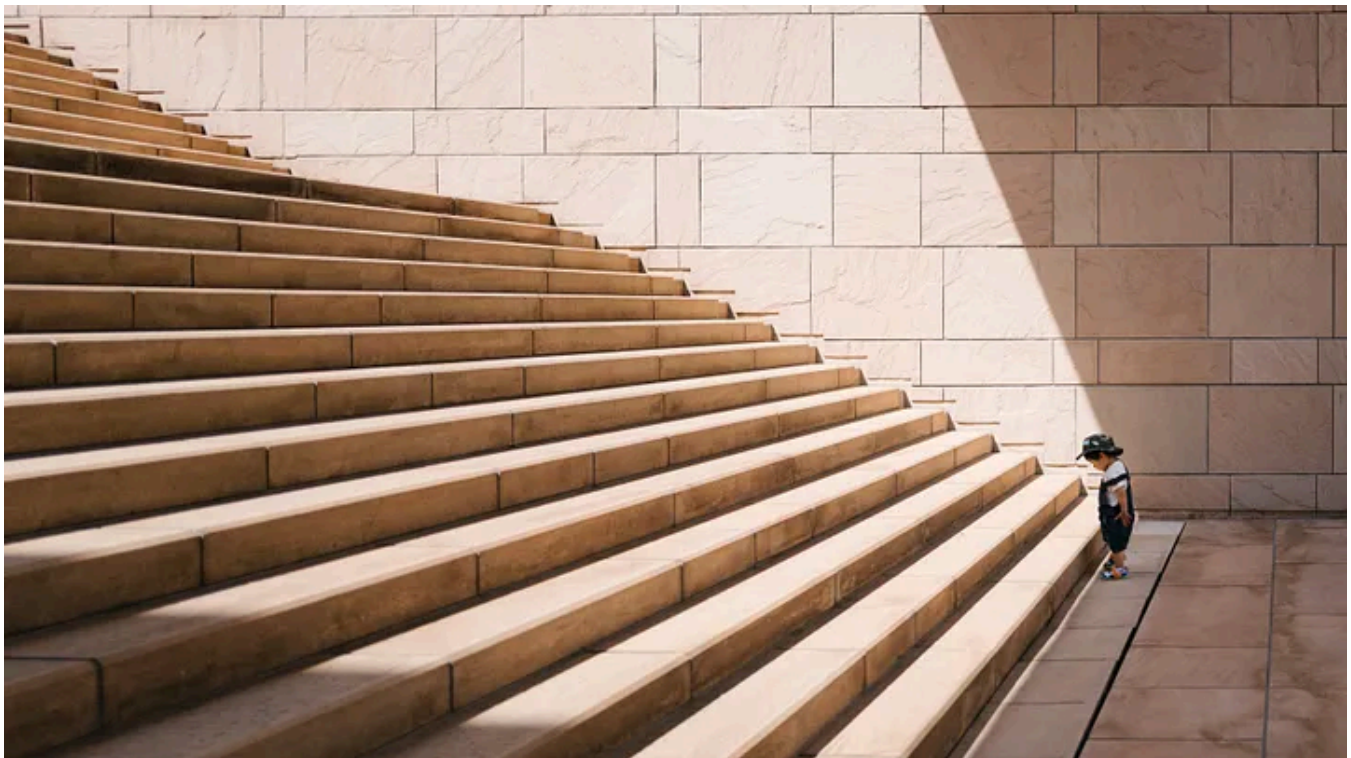
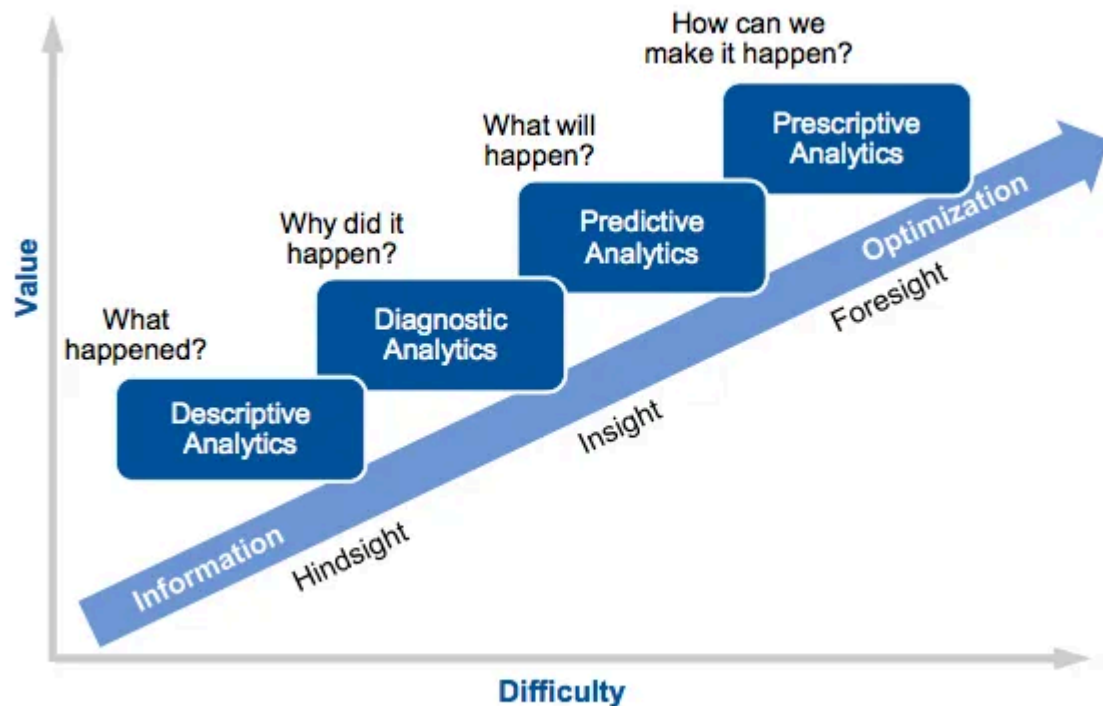


Photo by [Jukan Tateisi](#) on [Unsplash](#)

Since 2012, we have heard that famous saying from Gartner that analytics, in a general way, have four levels before we reach the state-of-art. Even though this article was published 9 years ago, it represents well (even too much) the reality of most companies.

The four steps, or the four major kinds of analytics can be defined as:

- Descriptive — What has happened?
- Diagnostic — Why did it happen?
- Predictive — What will happen?
- Prescriptive — How can we make it happen?



Source: Gartner (March 2012)

The famous Gartner ascendancy model

Descriptive Analytics

We can easily understand the first two since its idea has been well spread across companies. We can say that **descriptive** analytics came within the **first BI generation** and made Excel spreadsheets famous. The main idea is that we we can describe (most obvious meaning) the numbers of our organization:

- *What is the monthly revenue?*
- *How much we spent on certain expenses?*
- *How are the inventory levels?*

It is the first step of understanding our data, working on information from the **past**.

Diagnostic Analytics

Once we know what happened in the past, the second step is **diagnostic** analytics, asking why it happened. This is solved by using the **second BI generation**, which we can correlate different data, coming from different sources. Some example questions are:

- *Why our revenue is always higher in the end of the year?*
- *Why do we spend so much in automotive expenses?*
- *Why is our inventory decreasing so quickly?*

This next level requires a little bit of curiosity by the user, and the will to solve problems. Here we focus on the **actual** problems and **why** they happen.

If we take a closer look, we could say that descriptive analytics looks back to the **past** while diagnostic tends to observe the **present**. By this logic, you shall be thinking that the next level should be to look upon the **future**. And guess what, you are right! What most people misunderstands is that **only** predictive analytics observes the future, but not prescriptive.

And then, after all this analytics maturity 101, we come back to the main point of this article.

What is the difference between predictive and prescriptive analytics?

The third and the fourth levels both were made famous by the advent of Data Science and Machine Learning. Do not mistake these terms by 'foreseeing' the future, this is still impossible even with the technology we have nowadays. What we data scientists do is **infer** (or predict) the most probable scenario, **based on historical data**. So, the real questions are **not** like:

- *What will be my revenue for the next 6 months?*
- *How much will we spend on automotive expenses next month?*
- *How much inventory should we have to not have backlog for the next weeks?*

But rather:

- *What will be my revenue for the next 6 months **based on the last two years**?*

- *How much will we spend on automotive expenses next month **based on** the historical usage of our fleet, brand of the vehicle and the numbers of open delivery orders?*
- *How much stock should we have to not have backlog for the next weeks **based on** RFM and number of campaigns made last month?*

Machine Learning has **always** to be based on historical data, we cannot infer something out of thin air. And also, we use the questions answered in the diagnostic phase as initial stage of the model development.

What is important to notice is the output of predictive analytics:

- *What will be my revenue for the next 6 months **based on** the last two years?*

Ans: \$ 14,242,924.52

- *How much will we spend on automotive expenses next month **based on** the historical usage of our fleet, brand of the vehicle and the numbers of open delivery orders?*

Ans: \$ 241,242.08

- *How much inventory should we have to not have backlog for the next weeks **based on** RFM and number of campaigns made last month?*

Ans: 6,236 items

And, we also get a model, which can be a mathematical equation, that we use to generate these numbers, for example:

*Revenue = 1152.94 + 32.98 * <number of products sold> + 790 * <number of campaigns made>*

For example, let us say we sold 35,743 products and made 15 campaigns. We get:

*Revenue = 1152.94 + 32.98 * 35,743 + 790 * 15*

Revenue = \$ 1,191,807.08

Of course, we use much more advanced techniques to create models, but that is the main idea of machine learning: creating an 'equation' based on historical data to predict the most **probable** future scenario. Focus on that word, **probable**.

And what about prescriptive analytics, we wasted up our timeline: past, present, and future. What is there left?

At prescriptive analytics we focus on **achieving** these probable numbers. The questions now are:

- *How do we achieve a revenue of \$ 14,242,924.52?*
- *How do we budget the automotive expenses at \$ 241,242.08?*
- *How do we guarantee we will not have backlog for certain item?*

And some of the answers could be:

- *How do we achieve a revenue of \$ 14,242,924.52?*
Ans: Reducing margin by 3% and increasing sales by 10%
- *How do we budget the automotive expenses at \$ 241,242.08?*
Ans: Use this route for vehicle 1, this route for vehicle 2...
- *How do we guarantee we will not have backlog for certain item?*
Ans: Use the 6,236 items as a baseline and add 10% more as safety margin. Also, buy these items with these suppliers: <list>

In summary, prescriptive analytics are focused on the **decision** and/or the **action**.

The main difference in predictive and prescriptive analytics is that, in predictive analytics, we have a machine **helping us to take decisions**, while in prescriptive analytics we will have the machine **telling us what to do** to achieve the numbers we got in predictive analytics. Whether we will use the machine recommendation or not will be a human decision.

But why do we need a predictive model to build a prescriptive one?

Time for some hands-on

Let us have an example. **Pricing** is a retail practice of defining a price that optimizes **profit** (not revenue). In the example, we will use a private database, and we will not focus on the code, but on the business case.

We already know from basic administration courses in the academy that:

$$\text{Profit} = \text{Revenue} - \text{Cost}$$

$$\text{Revenue} = \text{Quantity} * \text{Price}$$

So,

$$\text{Profit} = \text{Quantity} * \text{Price} - \text{Cost}$$

We can estimate a simple demand model, where we target the **quantity** based on the **price** only. And we know that the higher the price, the fewer products are sold.

Let us use some data about a certain product in retail:

		Price	Quantity
Year	Month		
2016	01	103.31	203.0
	02	108.66	184.0
	03	112.57	211.0
	04	107.40	213.0
	05	117.23	241.0
Open in app ↗			



Search



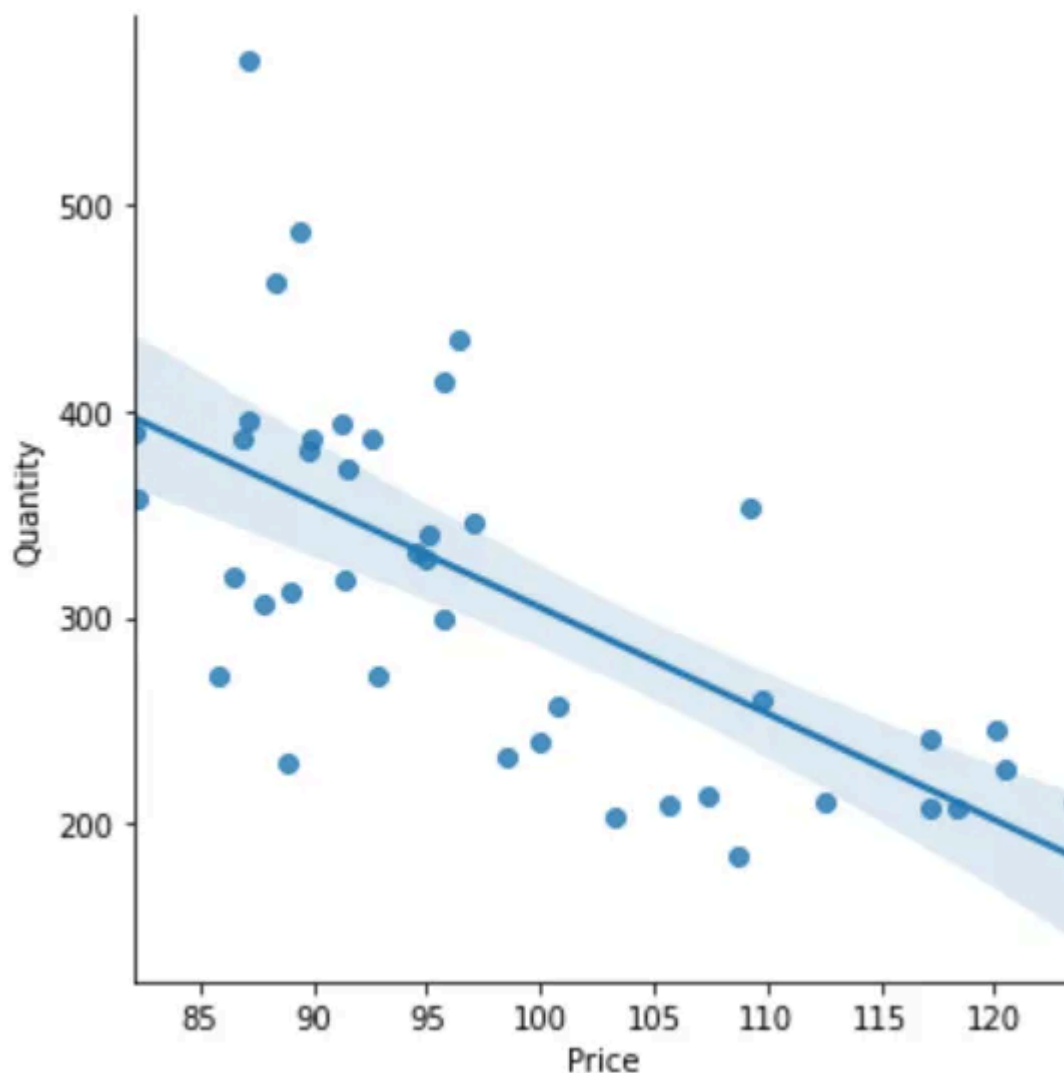
08 120.09 246.0

09 120.49 226.0

10 117.14 208.0

Monthly sales of a single random product

If we plot this data, we can see a downward pattern. As we expected, the more we increase the price, the less products we sell.



Linear Model for Price x Quantity (made with Seaborn)

From the Demand Theory, we should expect the line equation ($y = ax + b$) like:

$$Quantity = -a * Price + b$$

Where the coefficients a and b will be defined by the statistical model. The minus sign in a reflects the downward trend.

We will use the simplest model available to us, a Linear Regression (using Ordinary Least Squares or OLS). Once we create the model, we get the following results:

OLS Regression Results

Dep. Variable:	Quantity	R-squared:	0.439			
Model:	OLS	Adj. R-squared:	0.425			
Method:	Least Squares	F-statistic:	31.28			
Date:	Tue, 13 Apr 2021	Prob (F-statistic):	1.76e-06			
Time:	14:22:35	Log-Likelihood:	-235.89			
No. Observations:	42	AIC:	475.8			
Df Residuals:	40	BIC:	479.3			
Df Model:	1					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
Intercept	818.7170	90.812	9.016	0.000	635.180	1002.254
Price	-5.1388	0.919	-5.593	0.000	-6.996	-3.282
Omnibus:	4.248	Durbin-Watson:	1.483			
Prob(Omnibus):	0.120	Jarque-Bera (JB):	3.031			
Skew:	0.602	Prob(JB):	0.220			
Kurtosis:	3.533	Cond. No.	853.			

Model created with python lib statsmodels

Even though there is a lot of information here, we will focus on the **coef** column, where we get the information for the Intercept and the Price. The rest of the information tell us about the model performance, which we will not cover here.

Our Quantity x Price model now becomes:

$$Quantity = 818.72 - 5.14 * Price$$

Of course, we have an error associated with this equation, we can notice on the Quantity x Price scatterplot a lighter blue area, that is the error. A huge part of our jobs is to minimize this error as much as possible, so we get more accurate results.

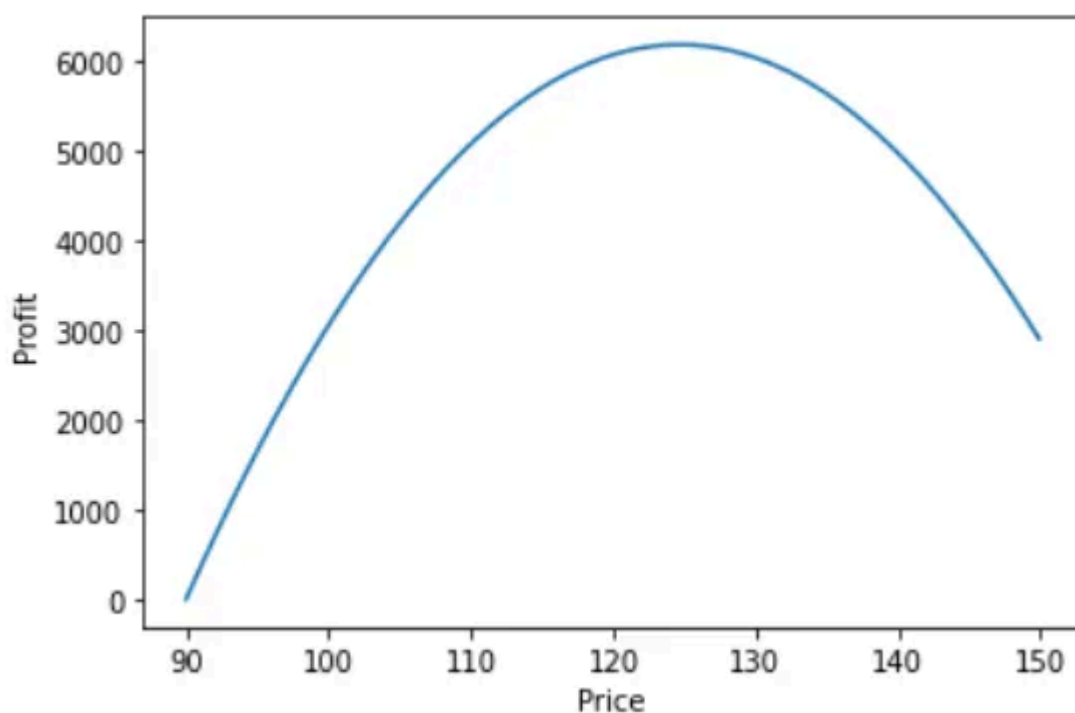
Anyways, the final equation we get is the equation that represents is most **probable** (here is that important word again) scenario that will occur.

Once we have this (predictive) model, we can substitute in our profit equation:

$$\text{Profit} = \text{Quantity} * \text{Price} - \text{Cost} = (818.72 - 5.14 * \text{Price}) * \text{Price} - \text{Cost}$$

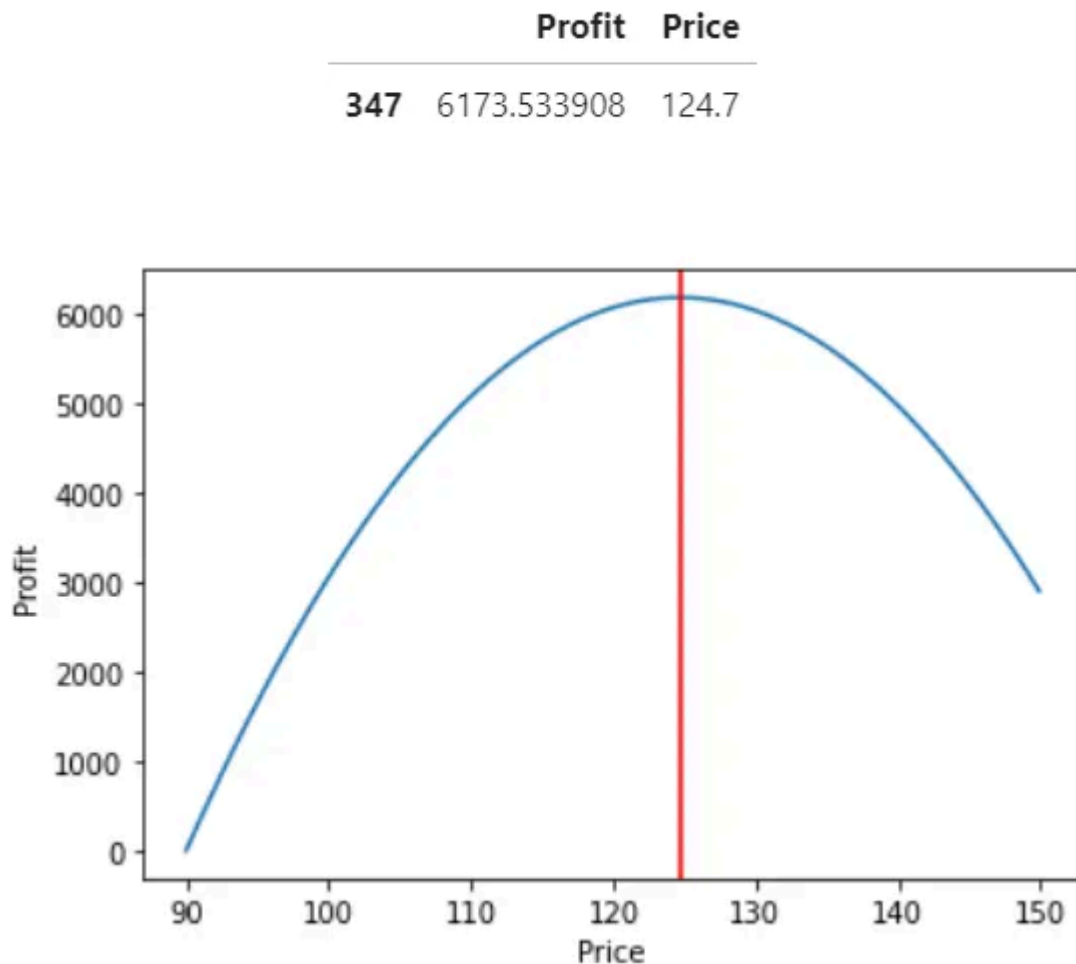
We notice that we have the price squared, and again coming back to our college classes, we know that these equations whose coefficients are powered to 2 (or second order equations) have a curve, which we can estimate the **maximum point** that represents the maximum profit we can get from a certain product.

For the cost, we will assume a fixed value, but as you may have noticed, we could also create a cost model based also on demand. But let us keep things simple, we will assume that the cost is \$ 90.00 per product. Plotting this equation we will get the following chart:



Profit curve

Where we can easily get the maximum profit by setting the price at \$124.70, and by that, ending of our **prescriptive** analysis.



Maximum value chart and value

At the end of the day, what we will have is a robot recommending your company a price for every single product you sell.

Wrap-up

I hope you enjoyed reading this article as I have writing it, and I hope it helps you to understand where you are, and where you want to be in your analytical journey. Notice that every analytical level is a step to the next one, so do not rush to get to the last phase without crossing the previous levels. Each one of the major kinds of analytics will give you the right information you need to ascend to the next one.

[Data Science](#)[Towards Data Science](#)[Analytics](#)[Machine Learning](#)[Prescriptive Analytics](#)

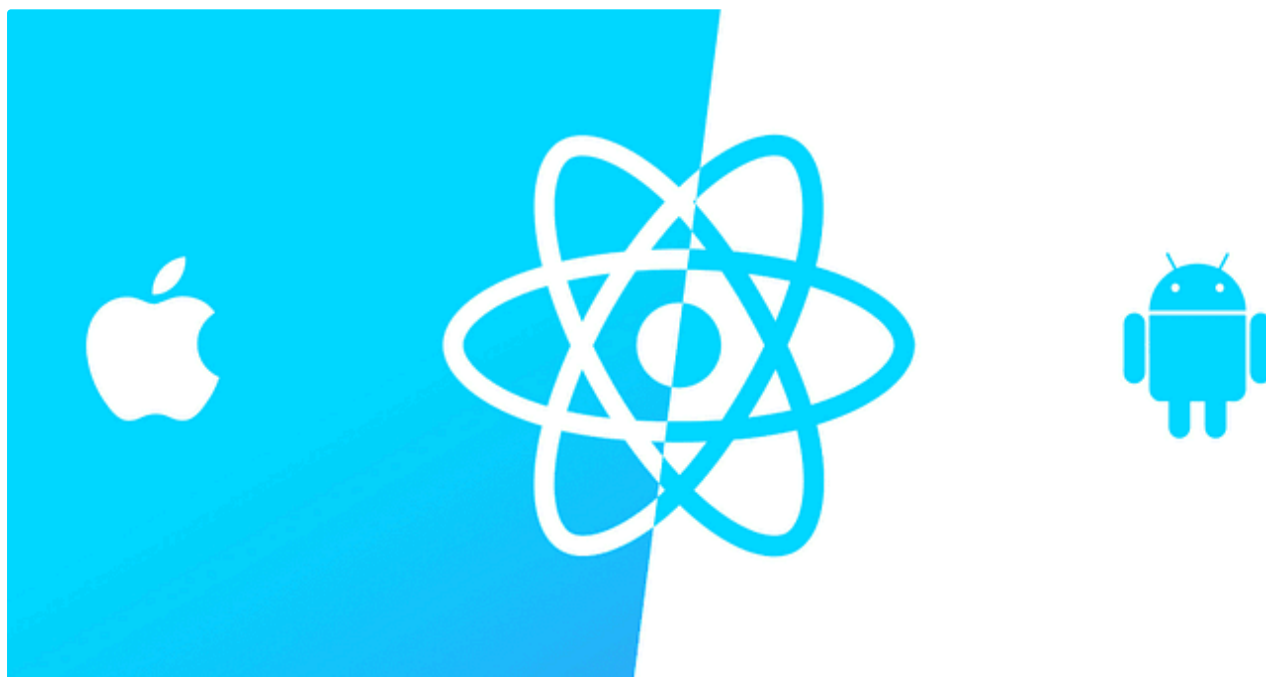
[Follow](#)

Written by Victor Almeida

9 Followers · Writer for Geek Culture

Technology enthusiast, a hell of a curious person, trying to make people understand complex things. Also, a Data Scientist by passion

More from Victor Almeida and Geek Culture




 Anshul Borawake in Geek Culture

React Native Generate APK—Debug and Release APK

Generate Debug and Release APK in React Native; Windows, iOS and Linux

3 min read · Apr 3, 2021

 1.7K  11



Masud Afsar in Geek Culture

How to install Node.js by NVM?

Install and manage multiple versions of Node.js with nvm

3 min read · Sep 14, 2021



146



2



Hasitha Subhashana in Geek Culture

Circuit Breaker Pattern (Design Patterns for Microservices)

In a distributed system we have no idea how other components would fail. Network issues could occur, components could fail or a router or a...

10 min read · Jun 12, 2021



618



5



Shu Ishida in Geek Culture

Installing Linux (Ubuntu 20.04) on an external portable SSD and pitfalls to be aware of

Carry your Ubuntu OS anywhere without being limited by hardware

11 min read · Apr 22, 2022



263



2



See all from Victor Almeida

See all from Geek Culture

Recommended from Medium



August Severn

Mastering Data Strategy: A Three-Part Video Guide

Previously posted: <https://capitoldataanalytics.com/mastering-data-strategy/>

2 min read · Mar 4, 2024





Somnath Singh in Level Up Coding

The Era of High-Paying Tech Jobs is Over

The Death of Tech Jobs.

🌟 · 14 min read · Apr 1, 2024



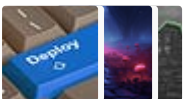
9.8K



253



Lists



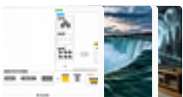
Predictive Modeling w/ Python

20 stories · 1135 saves



Practical Guides to Machine Learning

10 stories · 1365 saves



Natural Language Processing

1411 stories · 908 saves



data science and AI

40 stories · 139 saves



Roland Nagy

Foot Traffic Analysis—Customer Behaviour Analysis with Offline and Online Method

Foot traffic based data analysis can be key to the success of any physics business. Understanding customer behaviour can help to achieve...

🌟 · 5 min read · Dec 8, 2023



3



Benoit Pimpaud

The Data Analyst Every CEO Wants

Data Analyst is probably the most underrated job in the data industry.

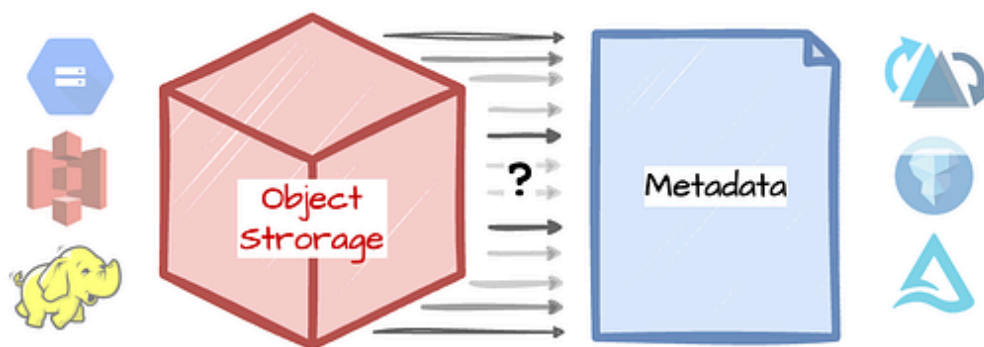
✦ · 7 min read · Apr 2, 2024



77



2



Do we need the Lakehouse architecture?



Vu Trinh in Data Engineer Things

Do We Need the Lakehouse Architecture?

When data lakes and data warehouses are not enough.

10 min read · Apr 20, 2024

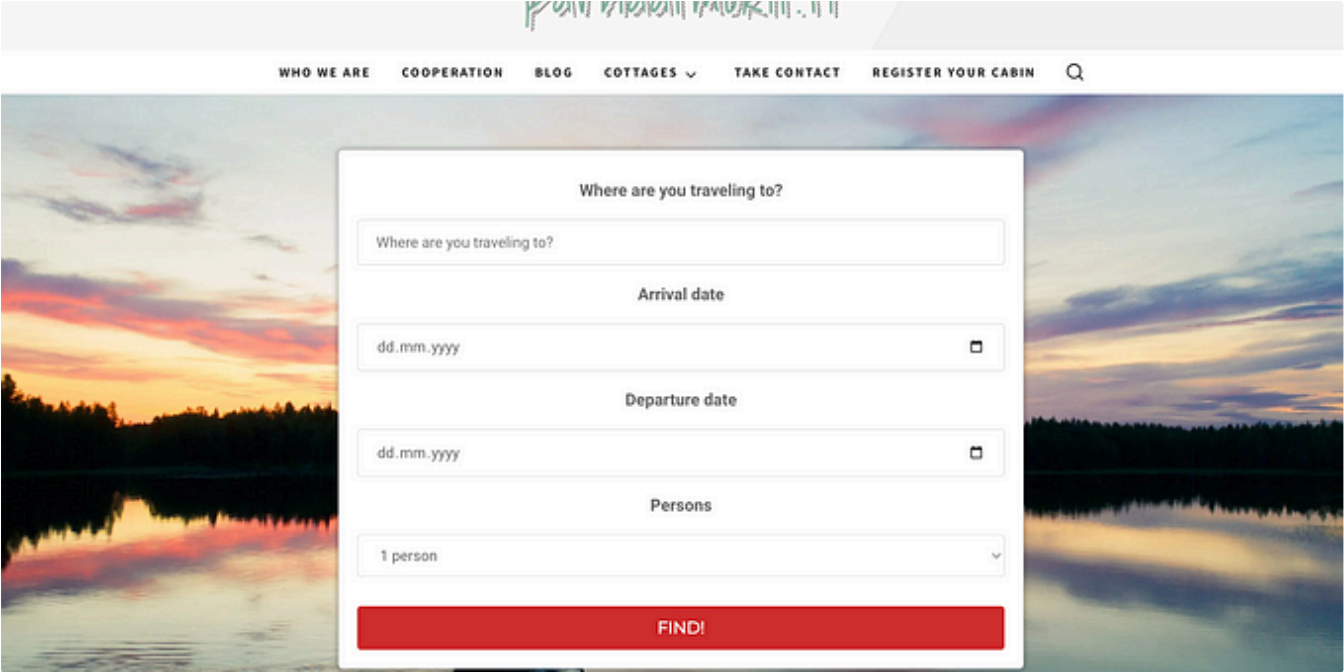



214



8





 Artturi Jalli


I Built an App in 6 Hours that Makes \$1,500/Mo

Copy my strategy!

🌟 · 3 min read · Jan 23, 2024

 17.5K

 188





See more recommendations