

Welcome back. You are signed in as **be.....@gmail.com**. Not you?

★ Get unlimited access to the best of Medium for less than \$1/week. [Become a member](#)



How to solve K-Means Algorithm Numerical?



Sujan Karna · [Follow](#)

2 min read · Dec 18, 2023



131



5



Welcome back. You are signed in as **be.....@gmail.com**.

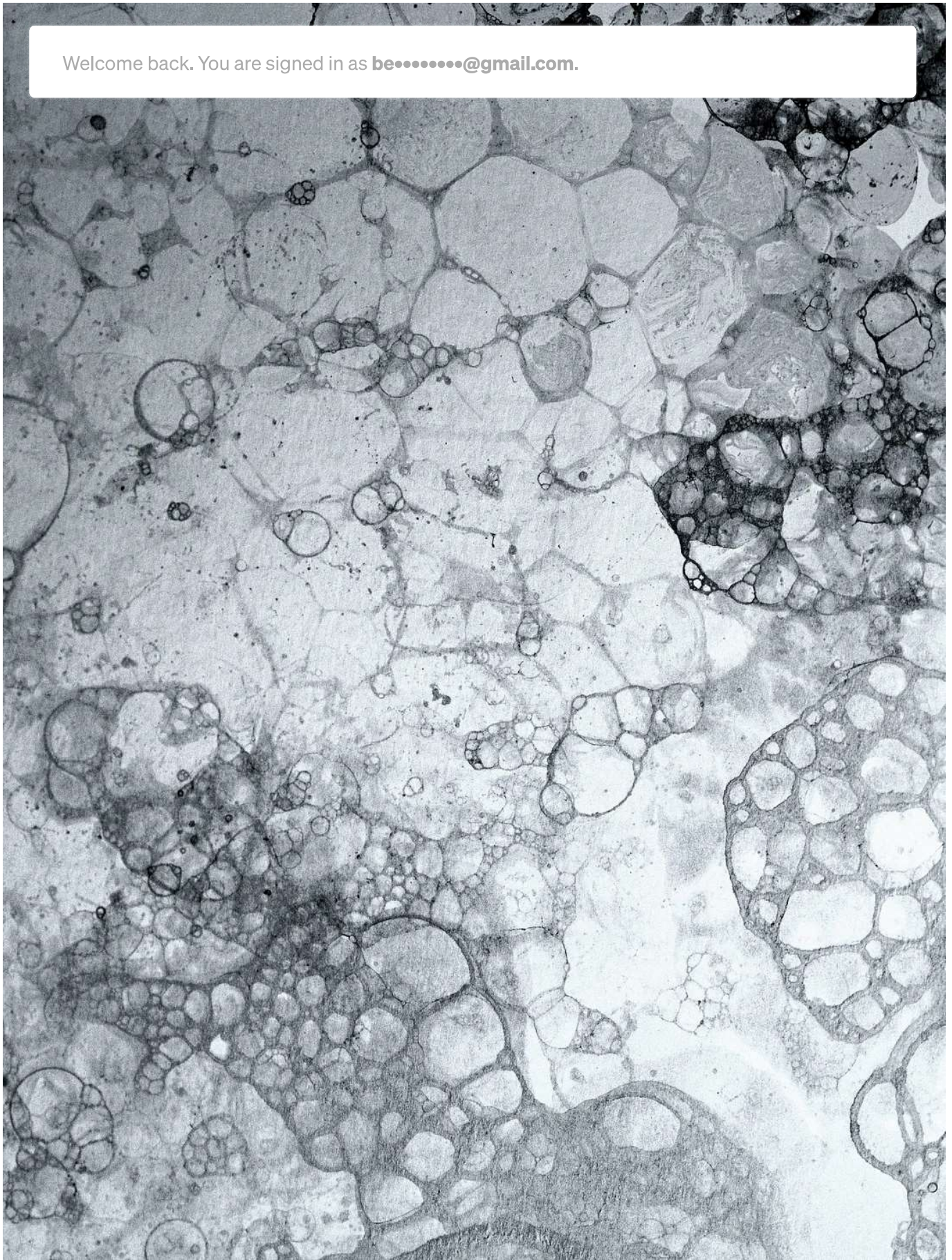


Photo by [Kier in Sight Archives](#) on [Unsplash](#)

Q. Apply $K(=2)$ -Means algorithm over the data (185 72) (170 56) (168 60)

(1. Welcome back. You are signed in as **be*****@gmail.com**.

ly

choose first two objects as initial centroids.

Solution:

Given, number of clusters to be created (K) = 2 say c_1 and c_2 ,

number of iterations = 2 and

The given data points can be represented in tabular form as:

Instance	X	Y
1	185	72
2	170	56
3	168	60
4	179	68
5	182	72
6	188	77

Data Points

also, first two objects as initial centroids:

Centroid for first cluster $c_1 = (185, 72)$

Centroid for second cluster $c_2 = (170, 56)$

Iteration 1: Now calculating similarity by using *Euclidean distance* measure

Open in app ↗

Medium

Search

Write



Welcome back. You are signed in as **be.....@gmail.com**.

$$d(c2, 3) = \sqrt{(170 - 168)^2 + (56 - 60)^2} = \sqrt{(2)^2 + (-4)^2} = \sqrt{4 + 16} = \sqrt{20}$$

Here, $d(c2, 3) < d(c1, 3)$

So, data point 3 belongs to c2.

$$d(c1, 4) = \sqrt{(185 - 179)^2 + (72 - 68)^2} = \sqrt{(6)^2 + (4)^2} = \sqrt{36 + 16} = \sqrt{52}$$

$$d(c2, 4) = \sqrt{(170 - 179)^2 + (56 - 68)^2} = \sqrt{(-9)^2 + (-12)^2} = \sqrt{81 + 144} = \sqrt{225}$$

Here, $d(c1, 4) < d(c2, 4)$

So, data point 4 belongs to c1.

$$d(c1, 5) = \sqrt{(185 - 182)^2 + (72 - 72)^2} = \sqrt{(3)^2 + (0)^2} = \sqrt{9}$$

$$d(c2, 5) = \sqrt{(170 - 182)^2 + (56 - 72)^2} = \sqrt{(-12)^2 + (-16)^2} = \sqrt{144 + 256} = \sqrt{400}$$

Here, $d(c1, 5) < d(c2, 5)$

So, data point 5 belongs to c1.

$$d(c1, 6) = \sqrt{(185 - 188)^2 + (72 - 77)^2} = \sqrt{(-3)^2 + (-5)^2} = \sqrt{9 + 25} = \sqrt{34}$$

$$d(c2, 6) = \sqrt{(170 - 188)^2 + (56 - 77)^2} = \sqrt{(-18)^2 + (-21)^2} = \sqrt{324 + 441} = \sqrt{765}$$

Here, $d(c1, 6) < d(c2, 6)$

So, data point 6 belongs to c1.

Euclidean distance calculation

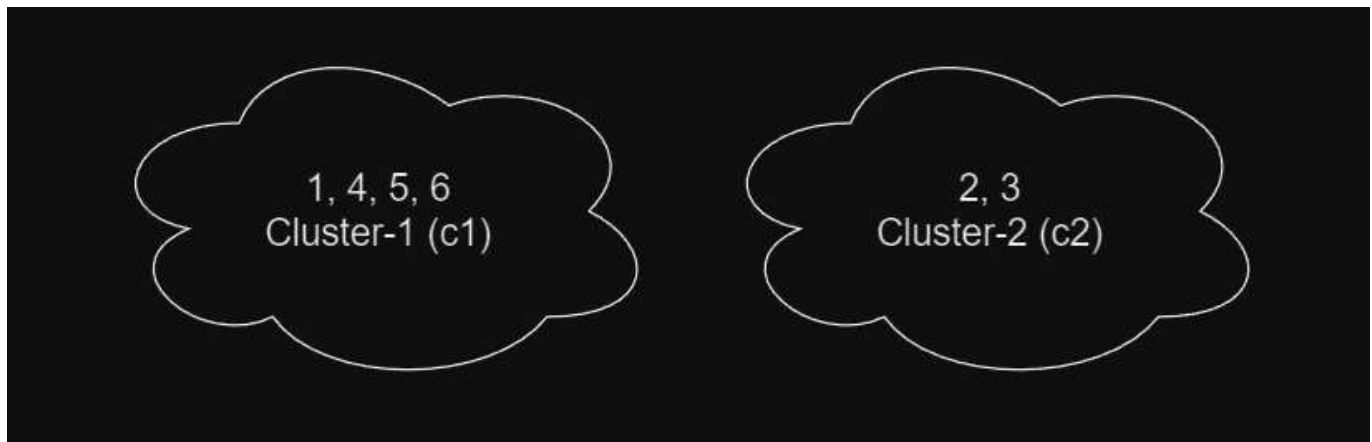
Representing above information in tabular form:

Instance	X	Y	Distance(C1)	Distance(C2)	Cluster
1	185	72			c1
2	170	56			c2
3	168	60	$\sqrt{433}$	$\sqrt{20}$	c2
4	179	68	$\sqrt{52}$	$\sqrt{225}$	c1
5	182	72	$\sqrt{9}$	$\sqrt{400}$	c1
6	188	77	$\sqrt{34}$	$\sqrt{765}$	c1

Distance of each data points from cluster centroids

The resulting cluster after first iteration is:

Welcome back. You are signed in as **be.....@gmail.com**.



Data points cluster

Iteration 2: Now calculating centroid for each cluster:

$$\text{Centroid for first cluster } c1 = \left(\frac{185+179+182+188}{4}, \frac{72+68+72+77}{4} \right) = (183.5, 72.25)$$

$$\text{Centroid for second cluster } c2 = \left(\frac{170+168}{2}, \frac{56+60}{2} \right) = (169, 58)$$

Calculating centroid as mean of data points

Now, again calculating similarity:

Welcome back. You are signed in as **be.....@gmail.com**.

$$d(c2, 1) = \sqrt{(169 - 185)^2 + (58 - 72)^2} = 21.2603$$

Here, $d(c1, 1) < d(c2, 1)$

So, data point 1 belongs to c1.

$$d(c1, 2) = \sqrt{(183.5 - 170)^2 + (72.25 - 56)^2} = 21.1261$$

$$d(c2, 2) = \sqrt{(169 - 170)^2 + (58 - 56)^2} = 2.2361$$

Here, $d(c2, 2) < d(c1, 2)$

So, data point 2 belongs to c2.

$$d(c1, 3) = \sqrt{(183.5 - 168)^2 + (72.25 - 60)^2} = 19.7563$$

$$d(c2, 3) = \sqrt{(169 - 168)^2 + (58 - 60)^2} = 2.2361$$

Here, $d(c2, 3) < d(c1, 3)$

So, data point 3 belongs to c2.

$$d(c1, 4) = \sqrt{(183.5 - 179)^2 + (72.25 - 68)^2} = 6.1897$$

$$d(c2, 4) = \sqrt{(169 - 179)^2 + (58 - 68)^2} = 14.1421$$

Here, $d(c1, 4) < d(c2, 4)$

So, data point 4 belongs to c1.

$$d(c1, 5) = \sqrt{(183.5 - 182)^2 + (72.25 - 72)^2} = 1.5207$$

$$d(c2, 5) = \sqrt{(169 - 182)^2 + (58 - 72)^2} = 19.1050$$

Here, $d(c1, 5) < d(c2, 5)$

So, data point 5 belongs to c1.

$$d(c1, 6) = \sqrt{(183.5 - 188)^2 + (72.25 - 77)^2} = 6.5431$$

$$d(c2, 6) = \sqrt{(169 - 188)^2 + (58 - 77)^2} = 26.8701$$

Here, $d(c1, 6) < d(c2, 6)$

So, data point 6 belongs to c1.

Distance calculation between data points and centroids

Representing above information in tabular form.

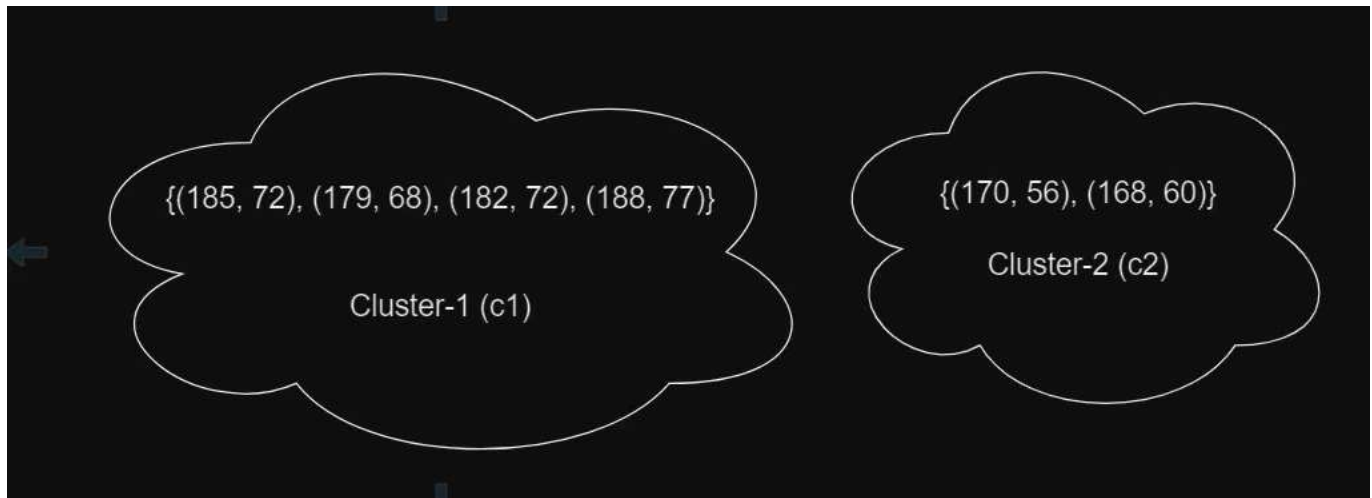
Instance	x	y	Distance(C1)	Distance(C2)	Cluster
----------	---	---	--------------	--------------	---------

Welcome back. You are signed in as **be.....@gmail.com**.

3	168	60	19.7563	2.2361	c2
4	179	68	6.1897	14.1421	c1
5	182	72	1.5207	19.105	c1
6	188	77	6.5431	26.8701	c1

Distance of each data points from cluster centroids

The resulting cluster after second iteration is:



Data points cluster

As we have already completed two iteration as asked by our question, the numerical ends here.

Since, the clustering doesn't change after second iteration, so terminate the iteration even if question doesn't say so.

Data Warehousing

Machine Learning

Data Mining

Artificial Intelligence

Clustering Algorithm

Welcome back. You are signed in as **be.....@gmail.com**.



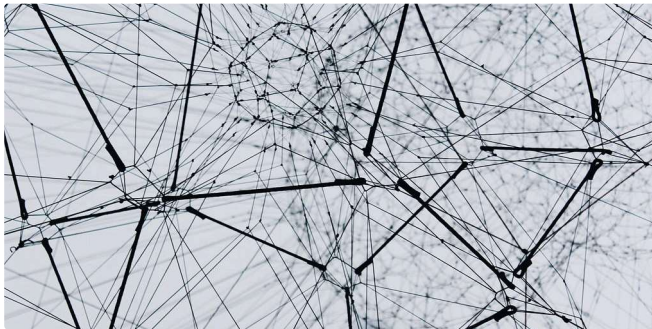
Written by Sujan Karna

Follow

33 Followers

 Passionate Web Developer & Data Science Enthusiast

More from Sujan Karna



 Sujan Karna

Back Propagation Algorithm - Numerical Solved

Q. Consider a multilayer feed-forward neural network given below. Let the learning rate b...

Dec 27, 2023

 41

 5



 Sujan Karna

Density Based- DBSCAN Numerical

Q. Given the points A(3, 7), B(4, 6), C(5, 5), D(6, 4), E(7, 3), F(6, 2), G(7, 2) and H(8, 4), Find the...

Dec 20, 2023

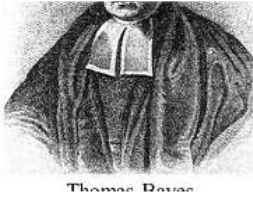
 9




Naive Bayes Classifier

Welcome back. You are signed in as **be.....@gmail.com**.


$$P(A|B) = \frac{P(B|A) P(A)}{P(B)}$$

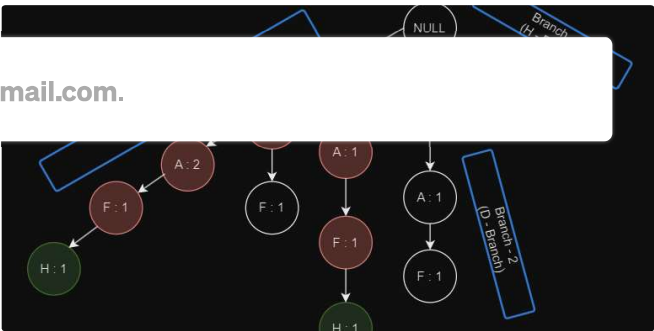


 Sujan Karna

How to solve Naive Bayesian Classification -Numerical?

Q. Consider the following data set.


Jan 7  62  1  



 Sujan Karna

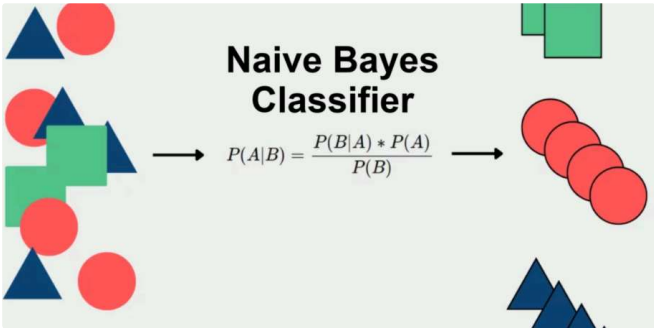
How to solve Frequent Pattern Mining-FP Growth Numerical?

Q. Generate the frequent pattern from the following data set using FP growth, where...

Jan 11  19  2  

See all from Sujan Karna

Recommended from Medium





Fraidoon Omarzai



Anurag



Welcome back. You are signed in as **be*****@gmail.com**.

This blog covers the fundamentals of Naive Bayes with clear explanations.

A Guide to Handling Categorical Variables in Machine Learning

Jul 20

👍 46



Apr 10



Lists



Predictive Modeling w/ Python

20 stories · 1548 saves



Practical Guides to Machine Learning

10 stories · 1877 saves



Natural Language Processing

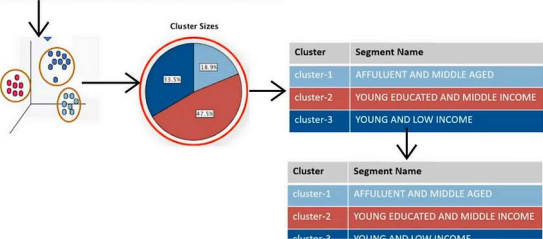
1715 stories · 1287 saves



The New Chatbots: ChatGPT, Bard, and Beyond

12 stories · 462 saves

2	47	1	26	500	4.582	8.218	NBA021	12.8	0
3	33	2	10	57	6.111	5.802	NBA013	20.9	1
4	29	2	4	19	0.681	0.516	NBA009	6.3	0
5	47	1	31	253	9.308	8.908	NBA008	7.2	0
6	40	1	23	81	0.998	7.831	NBA016	10.9	1
7	38	2	4	56	0.442	0.454	NBA013	1.6	0
8	42	3	0	64	0.279	3.945	NBA009	6.6	0
9	26	1	5	18	0.375	2.235	NBA006	15.5	1



Avicsebooks

ML Part 5: Clustering

Clustering is a type of unsupervised machine learning technique used for grouping similar...



Ansa Baby in Tech & TensorFlow

Understanding the Mathematical Concepts Behind the K-Nearest...

A Complete Beginner's Guide to the Math Behind KNN

Apr 17

👍 5



Sep 15

👍 53

💬 2



Sujan Karna

Software Development Engineer

• Developed Amazon checkout and payment services to handle traffic of 10 Million daily global transactions

Mar. 2020 – May 2021

Welcome back. You are signed in as **be.....@gmail.com**.

NinjaPrep.io (React)

- Platform to offer coding problem practice with built in code editor and written + video solutions in React
- Utilized Nginx to reverse proxy IP address on Digital Ocean hosts
- Developed using Styled-Components for 95% CSS styling to ensure proper CSS scoping
- Implemented Docker with Seccomp to safely run user submitted code with < 2.2s runtime

HeatMap (JavaScript)

- Visualized Google Takeout location data of location history using Google Maps API and Google Maps heatmap code with React
- Included local file system storage to reliably handle 5mb of location history data
- Implemented Express to include routing between pages and jQuery to parse Google Map and implement heatmap overlay



Alexander Nguyen in Level Up Coding

The resume that got a software engineer a \$300,000 job at Google.

1-page. Well-formatted.



Jun 1



22K



431



Oct 2, 2022



70



Thomas Lewandowski

Master the Classic Question: “What Does a Dip in the Metrics Mean?”

How I answer this common data analytics and product execution interview question. With I...

[See more recommendations](#)