## 一杯柠檬水的启蒙 A FResh Squeeze on DATA

用数据解决问题 PROBLEM SOLVING WITH DATA

•

0

0

0



# 一杯柠檬水的启蒙 A FRESH SQUEEZE ON DATA

#### 用数据解决问题 PROBLEM SOLVING WITH DATA



在晴朗的一天,克拉拉跑回家兴奋地说:

这真是一个好主意!如 果你们需要帮助就告诉 我。



当然啦,你可是很擅 长做广告牌呢!

**好消息!**我想要开一个柠

檬水铺,然后把卖柠檬水

的钱都捐给医院!



第二天,克拉拉和亚历克斯起得很早。他们在自己 家门前布置一张桌子和广告牌。并摆上了一大罐 柠檬水,和一叠杯子。

两个小时过去了,都没有人来买柠檬水。亚历克斯 伤心地看着姐姐,说:

**村宁** 栈蒙 7/×

<u>.</u>.;

6

# 为什么我们什么都没卖出去呀,克拉拉?

克拉拉开始思考她可能在计划中遗漏了什么。为了让更多的人来光顾他们的柠檬水铺,她可以做些什么不同的事情呢?

突然,克拉拉有了一个主意。



克拉拉思考着她的妈妈是如何解决工作中的问题的。克拉拉 的妈妈是一名数据科学家。她在一家使用数据解决各种问题 的公司工作。

22



也许**数据**可以帮助 我们解决这个问 题!

#### 什么是数据,它能 怎么帮助我们呢?

1

数据就是信息。它就在我们身边。我们每 天都在创造数据。我们身边有很多像妈 妈一样的人会通过数据解决各种问题。

> 好吧,但是数据怎么帮我们 卖掉更多的柠檬水呢?

数据可以告诉我们人们想在我们的 柠檬水铺上购买什么。我们现在还没 有任何数据。我们可以通过提问来收 集数据。

7

些东西。



#### 我们问什么问题呢?

 $\mathbf{C}$ 

让我们问问我们的朋友们,看 看他们想在柠檬水铺上买哪

克拉拉和亚历克斯问他们的朋友他们会在柠檬水铺买什么。

克拉拉和亚历克斯找到他们的朋友卡洛斯、伊莎贝拉、和伊森,并与他们交谈。 卡洛斯说他想要草莓柠檬水。伊莎贝拉想要为她的笔记本买些贴纸。伊森想要 最普通的柠檬水和一些可以用来装扮他的花园的彩绘石头。





#### 克拉拉和亚力克斯的柠檬水铺重新开始营业啦。这次为了吸引更多的 游客光顾,他们准备了柠檬水、贴纸、和彩绘的石头。

他们等了又等,但只卖出去了两杯柠檬水。

#### 克拉拉和亚历克斯回家到处找贴纸,并且画了很多石头。



威廉姆斯先生是他们妈妈的同事,他看到他们很伤心地待在柠檬水铺旁边。

你们为什么难过呀?

1

再跟我说说,你们都问了谁?

你好,威廉姆斯先生。我们的东西 卖得不好。我们询问了我们的朋友 该卖些什么,但好像并没有什么 用。

你们有问过大人吗?

#### 我们问了我们的三个好朋友。

我问了我在学校的好朋友 卡洛斯!

我们没有问过大人。我以为可 以从任何人身上获得数据。

数据确实可以来自任何人。但是要做出好的预 测或猜测会发生什么,我们必须与很多人交谈!

#### 大人想要的可能与孩子可能想要的不一样。每个 人都会有不同的想法,这取决于他们喜欢什么。

不能只问我们的朋友吗?

该找一个有更多人经过的地方。

 $\bigcirc \neg \bigcirc$ 

你说的对。我们也应该问问其他人的想法。

除此之外我还有一个建议,就是你们应

6

我们现在知道该怎么做了!谢谢 你,威廉姆斯先生!



邻居们给克拉拉和亚历克斯提供了许多关于他们想要买什么的不同答案,包括饼干、纸杯蛋糕、柠檬水、贴纸、彩绘石头和溜溜球!这次没有人要草莓柠檬水了。



抑品 柠檬水 贴纸 彩绘石头 曲奇饼干 纸杯蛋糕 溜溜球

#### 人们最想要的三样东西是什么?



看来邻居们最想要鲜榨柠檬水、饼干和溜溜 球。我们就在柠檬水铺卖这些东西吧!



1:1

#### 我们家附近好像很少有人会经过,叫上我们的 朋友,一起去寻找有更多人经过的地方吧。

克拉拉和亚历克斯叫来了诺亚、卢卡斯和米娅,他们都很 高兴加入柠檬水小队!

诺亚住在杂货店旁边, 卢卡斯住在公园对面, 米娅则住在 学校附近。

6

0

•

让我们分别数一数一个小 时内会有多少人经过我们 各自的家。

6

eserence (

他们一起数了经过他们到人意味着更多的顾客。



他们一起数了经过他们家的人数。孩子们知道,更多的路

# 地点 诺亚的家 (杂货店) 35 卢卡斯的家 (公园) 56 米娅的家 (学校) 21

让我们看看哪里是设立柠 檬水摊的最佳地点。

#### 你能把图表中的地点链接到正确的位置吗?

哪个地点经过的人最多?





2 July



柠檬水小队迎来了好多顾客,每个人都忙极了!







每个人都为自己做的好事感到自豪。柠檬水小队现在了解到数据 科学的力量。因此他们踏上了使用数据将世界变得更美好的旅程!

治病。

第二天,孩子们给了医院准备了一个大信封,里面放着一 封信和他们筹集到的所有捐款。

请用这笔钱 帮助更多的人 ●一 柠檬水小队

# 你能告诉我们更 多关于数据的知 识吗?



克拉拉,我喜欢你和亚力克斯利用少量的数 据来安排和布置柠檬水铺的方法。

数据科学家们会使用大量数据来解决更大的问题,例如自动驾驶汽车和海洋污染。







#### 计算机可以在没有标记数据的情况下学习吗?

是的!我们还可以教计算机从规律中学习。我和你只是看一眼 柠檬和草莓,就知道它们是不同种类的水果,并不需要任何人 告诉我们他们不一样。



输入



输出

6 6

如果我们给计算机大量柠檬和草莓的图像,并要求计算机将这 些图像分成两类,计算机就可以查看图像提供的颜色、形状、或 其他特征。





我们为计算机提供足够的信息——输入,来让它给出我 们想要的答案——**输出**。

当我们利用规律时,我们不会为计算机标记数据,但是我们给计 算机提供数百万张图像以便让计算机查看的事物的特征。





特征



特征









我们还学到,我们不光要问小孩, 也需要像大人询问有关我们柠檬 水铺的问题。

如果你们只包含一部分人的想法,那么你们收集的 数据可能存在偏见。偏见意味着数据无法准确反映 所有人的想法。我们需要听到更多人的声音,因为我 们将要做出的决定也可能会影响到他们。

孩子们对食物的口味可能会产生什么样的偏见?





一个好医生吗?



这些数据对我们的日常生活 有何帮助?

数据可以通过很多不同的方 式帮助我们生活的方方面 面。数据科学家的一部分工 作就是收集所需的数据以用 来解决问题。

我喜欢音乐。数据可以在音 乐方面帮助我吗?

可以的。通过从你的播放列表中收集 的数据,我们可以帮助计算机为你推 荐你可能会喜欢的下一首歌曲。

是的。利用从患者那里收集的数据,研究人员 可以研究并寻找新疾病的治疗方法。除了音乐 和医学,数据科学还可以通过预测来帮助人们 做出更好的决定。

例如,我们用数据预测人们下周可能在杂货店 购买的牛奶数量,这样商店就知道应当储存多 少牛奶。再比如说,我们用数据预测下一小时 的交通情况,这样就能帮你和其他学生找到最 快的上学路线。

我长大后想成为一名医生。数据可以帮助我做





# 现实世界中 的问题

数据科学家们使用数据和类似的方 法来解决会影响到更多人的大问题。

尝试完成下面的小任务。利用数据来 回答问题。



#### 问题: 游戏公司应该开发哪种类型的游戏?



问题:





#### 数据如何让我的快递可以更快到达?

#### 问题: 如何利用数据来识别一只狗的品种?



提示:每个品种有哪些特点呢?

回答:



问题: 数据科学可以帮助治疗疾病吗?

回答:



#### 问题: 我们如何决定让宇航员将哪些食物带 到太空?

	<b>质量</b> (食物占据多少空间)	<b>卡路里</b> (食物可以提供多少能量)	
<b>食物A</b>	100	200	
<b>食物B</b>	50	150	
②     ◆	200	250	*
回答:			
		*	

题吗:

#### 我的问题是:

#### 你能想到一个你想利用数据回答的问



# 数据科学家们使用数据



大家好,

上学习的过程。

在我们的世界中,了解数据从哪里来,以及数据如何被用来 做好的预测和决定,是十分重要的。数据可以帮助我们找到 为人类、我们的自然资源、和我们的社区最好的解决问题的 方案。通过分析数据,我们可以解决很多问题。数据带来的 可能性是无限的——就像未来的你们一样!

数据能力的道路。

"Scott Aronson 两个小男孩的骄傲父亲& Cloudera 首席运营官"

非常感谢你们来读这本书!我希望你们跟我,还有我的两个 儿子Flynn和 Jedd一样,都很享受从克拉拉和亚历克斯身

"感谢你们这么早就开始踏上这段欣赏数据的价值并培养

你在路上——继续前进!"

词汇表

偏见(数据偏见):现有数据不能公平地代表被研究的人或对象。

分类:我们将事物归入的不同的组。

数据:你知道的事实或事物的集合,例如数字、文字、图像或仅仅是对事物的描述。

输入:提供给计算机的信息。在计算机科学中,输入通常是你想要分析的数据。

标记数据:带有一个或多个标签的信息,例如一张带有"猫"标签的猫的图片。

**输出:**计算机生成的数据。在机器学习中,输出是由模型生成的预测。通常,当 输入改变时,输出也会改变。

预测:对于将来会发生的事情进行猜测。

自动驾驶汽车:一种能够感知周围环境并自行移动的车辆。

未标记数据:没有标签的数据。例如,一张不带有"猫"标签的猫的图片。

www.freshsqueezekids.com

www.readyai.org/a-fresh-squeeze-on-data

贡献者: Annabel Hasty, Dave Tourezky, Ethan Chen, Joel Wilson, Juno Schaser, Kelli Lawless, Madge Miller, Melanie Beck, Roozbeh Aliabadi, Santiago Giraldo, Sushil Thomas, Zhinoos Mobasherfar

作者和插画师:Shanshan Jin



## A FRESH SQUEEZE ON DATA

#### PROBLEM SOLVING WITH DATA



One sunny day, Clara runs home and tells her family:

That's a nice idea! Let me know if you need any help!

**BIG NEWS!** I want to raise money for the hospital. I'm going to open a lemonade stand!

Can I help?

Sure, you're great at making signs!



The next day, Clara and Alex wake up early. They set up a table with a sign in front of their house. They lay out a jug of fresh lemonade and a stack of cups.

No one bought lemonade from them for two hours. Alex looks sadly at his sister and says,



-Emongde

# Why didn't we sell anything, Clara?

Clara begins to think about what she might have missed in her planning. What could she have done differently to bring more people to their lemonade stand?

Suddenly, Clara has an idea.

Clara thinks about how her mom solves problems at work. Her mom is a data scientist and works at a company that uses data to solve all kinds of problems.



Maybe **data** can help us solve our problem!

What is data and how can it help?

> Data is information. It's all around us. We create data every day and people like mom use it to solve problems.

> > Okay, but how can data help us sell more lemonade?

Data can tell us what people want to buy at our lemonade stand. We don't have any data yet. We have to collect data by asking questions.



Clara and Alex ask their friends what they would buy at a lemonade stand.

Clara and Alex talk to their friends Carlos, Isabella, and Ethan. Carlos wants strawberry lemonade. Isabella wants stickers for her notebooks. Ethan wants regular lemonade and some painted rocks for his garden.



#### Clara and Alex go back home to collect stickers and paint rocks.



Hoping to get more visitors to their lemonade stand, they reopen with lemonade, stickers, and painted rocks.

They wait and wait but only sell two cups of regular lemonade.



Mr. Williams, their mom's coworker, sees them looking sad at their stand.

Why are you all so sad?

Hi, Mr. Williams. Our things are not selling well. We asked our friends what we should sell, but it didn't help. Tell me more. Who did you ask?

Did you ask any adults?

We didn't ask the grown-ups. I thought data could come from anyone.

We asked three of our friends.

I asked Carlos, my friend from school!

It's true that data can come from anyone. But to make good predictions, or guesses about what will happen, we have to talk to a lot of people!

What an adult might want is different from what a kid might want. Everyone will have different ideas depending on what they like.

0

Not just our friends?

Another piece of advice is that you want to find a place that has more people walking by.

0

 $\bigcirc$ 

You're right. We should ask other people, too.

We know what to do now! Thanks, Mr. Williams! Clara and Alex begin asking people they meet in the neighborhood, adults and kids.





1.1



The neighbors give Clara and Alex lots of different answers about what they'd like to buy, including cookies, cupcakes, lemonade, stickers, painted rocks, and yo-yos! No other people ask for strawberry lemonade.

Wish List
Lemonade
Stickers
Lemonade, painted
Lemonade, painted
cookies & cupcake
cookies & cupcade
yo-yos, Lemonade • Yo-Yos Lemonade & cupcakes! cookies! Lemonade & vookies.

Item Lemonade stickers painted rocks cookies cupcakes Y0-405

#### What three things do people want most?



Our neighbors want freshly squeezed lemonade, cookies, and yo-yo's the most. Let's sell these things at our lemonade stand!

11:

Not many people walk past our house. Let's gather our friends and look for places that have more people walking by. Clara and Alex meet up with Noah, Lucas, and Mia, who are all excited to join the Lemonade Crew!

Noah lives by the grocery store, Lucas lives across from the park, and Mia lives near the school.

6

0

Let's count how many people go by our houses in one hour.

> Together, they count the number of people walking by their houses. The kids know that more people going by means more customers.



Which location had the most people passing by?

Can you link the correct place to the chart?





stand. It is right across the street from the park where the most people are walking by.



The Lemonade Crew has so many customers, and everyone is busy!



# Lemonade



Everyone is so proud of the good things they have done. The Lemonade Crew now understands the power of data science. They are on their journey to make the world better using data!

The next day, the kids give the hospital a big envelope with a letter and all the money they raised.

Please use this money to help people get better.

✓ - the Lemonade Crew

# **CAN YOU TELL US MORE ABOUT DATA?**



Clara, I like the way that you and Alex used a small amount of data to set up the lemonade stand.

Data scientists use a lot of data to solve bigger problems like self-driving cars and ocean pollution.















Can computers learn without labeled data?

Yes! We can also teach a computer to learn from patterns. You and I can look at lemons and strawberries and know they're different kinds of fruit without anyone telling us.







OUTPUT

We give computers enough information or **input** to provide the answers, or **output**, we want.

When we use patterns, we won't have the data labeled for the computers, but we have millions of images that can provide features for computers to look at.

If we give the computer a lot of images of lemons and strawberries and ask the computer to separate these images into two categories, the computer could look at the color, shape, or other features that the images provide.









FEATURE







We ask people questions and watch what they do. It's just like what you did with the lemonade stand, Clara and Alex. You asked what people wanted to buy and then saw what they actually bought.



We also learned that we need to ask both kids and adults for information for our lemonade stand.

If you only include one group of people's ideas, the data that you collect could be biased. **Bias** means that the data doesn't give a fair picture of what all the people think. We need to hear more people's voices because the decisions we will make might affect them as well.

How might kids' taste in food be biased?



I want to be a doctor when I grow up. Can data help me with medicine?

Yes. With the data collected from patients, researchers can study and find cures for new diseases. Besides music and medicine, data science helps people make better decisions.

For example, we **predict** the amount of milk people might buy next week at the grocery store, so the store knows how much to stock. We predict the traffic in the next hour and find the fastest way to get you to your school in the morning.



How does this data help us in our daily lives?

There are so many ways that data helps every aspect of our lives. Part of the data scientist's job is to collect the desired data to solve problems.

I love music. Can data help me with music?

Yes. With data collected from your playlist, we can help computers recommend the next song that you might like.





# **REAL-WORLD PROBLEMS**

Data scientists use data and similar methods to solve larger problems that affect more people.

Try the activities below. Use data to answer the questions.

#### Question: Which type of video game should a company create?



Question: How does data l arrive faster?



Answer:

#### How does data help my deliveries

#### **Question**: How can I use data to identify a dog's breed?



Hint: What are some of the features of each breed?

#### Answer:



**Question**:



Answer:

#### Can data science help cure diseases?

#### **Question:** How do we decide what food to take to space?

	MASS (how much space does the food take up)	calories (how much energy can it provide)	
FOOD A	100	200	
FOOD B	50	150	
FOOD C	200	250	
Answer	:		

My question is:

#### Can you think of one question that you want to answer with data:



# A Letter to the Reader



Hello Everyone,

Thank you so much for reading this book! My sons, Flynn and Jedd, and I really enjoyed learning from Clara and Alex and hope you did too.

In our world, it's very important that we understand where data comes from and how it can be used to make good predictions and decisions. Data can help us find the best solutions for people, our natural resources, and our communities. There are a lot of problems we can solve by analyzing data. The possibilities are limitless – just as they are for you!

Thanks for starting early on your path to appreciating the value of data and developing your data literacy. You're on your way – keep it going!

Scott Aronson

Proud father of two young boys & Chief Operating Officer at Cloudera

## GLOSSARY

Bias (Data bias): The available data does not fairly represent the people or objects in the study.

Classes: The different groups that we sort things into.

**Data:** A collection of facts or things you know, such as numbers, words, images, or just descriptions of things.

**Input:** Information given to the computer. In computer science, inputs are usually data you wish to analyze.

Labeled data: Information that has been tagged with one or more labels, such as a picture of a cat with the label "cat".

**Output:** Data generated by a computer. In machine learning, outputs are predictions generated by the model. Usually, output changes when the input is changed.

**Predict:** To guess about what you think will happen.

**Self-driving cars (autonomous car):** A vehicle that is capable of sensing its environment and moving by itself.

**Unlabeled data:** Data that comes with no tag. For example, a picture of a cat without the label "cat".

www.freshsqueezekids.com www.readyai.org/a-fresh-squeeze-on-data

Contributors: Annabel Hasty, Dave Tourezky, Ethan Chen, Joel Wilson, Juno Schaser, Kelli Lawless, Madge Miller, Melanie Beck, Roozbeh Aliabadi, Santiago Giraldo, Sushil Thomas Author and Illustrator: Shanshan Jin A Book By

# CLOUDERA

h

22

In partnership with Ready

É.

