



# Universal Models

Forecasting, Personal Data Protection, Sentiment Analysis, and Beyond

*AI Labs Team*

*July 2024*

# What Is A Model?

A model is an equation

## ***Simple equation:***

- $a + b = c$
- 1 apple + 3 apples = 4 apples
- \$300 + \$400 = \$700

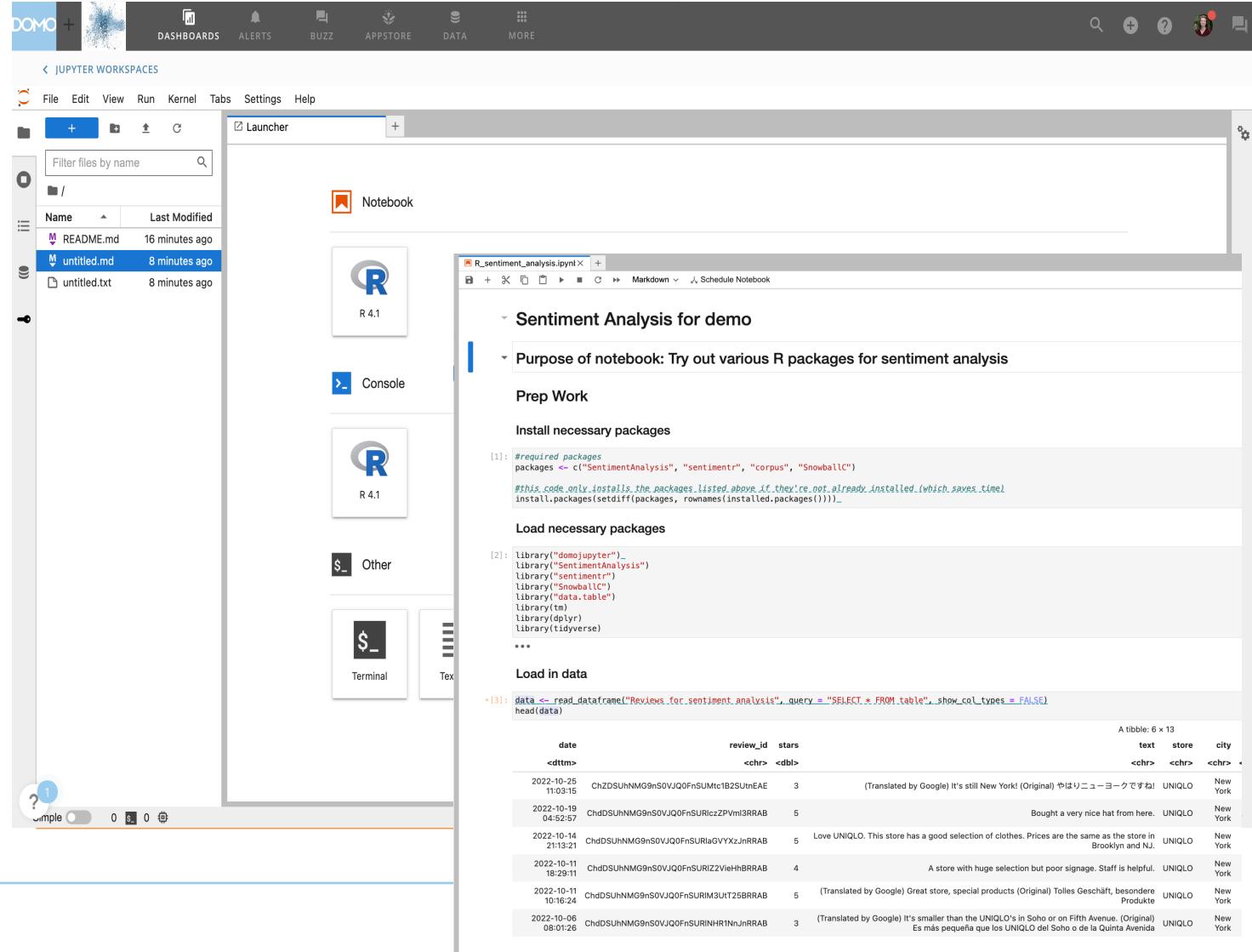
## ***Forecasting model:***

- $c + \phi_1 y_{dt-1} + \phi_p y_{dt-p} + \dots + \theta_1 e_{t-1} + \theta_q e_{t-q} + e_t = Y_t$
- $\$5k + (0.6 \cdot \$10K) + (0.3 \cdot \$8K) + (0.4 \cdot \$2K) + (0.2 \cdot \$1.5K) + \$10K = \$15.5K$  revenue next week

# Custom Models in Domo

## CUSTOM MODELS

- Models that have been specifically built based on your data & use case
- Built using Jupyter Workspaces, AutoML, or Magic ETL Scripting Tiles
- Requires advanced statistical knowledge &/or coding skills
- Takes time to develop & maintain
- Better performance & accuracy



The screenshot shows the Domo Jupyter Workspaces interface. At the top, there's a navigation bar with links for Dashboards, Alerts, Buzz, Appstore, Data, and More. Below that is a search bar and a user profile icon.

The main area has a left sidebar labeled "JUPYTER WORKSPACES" with a "Launcher" tab selected. It shows a file browser with files like README.md, untitled.md, and untitled.txt. There are also icons for Notebook, Console, Other, Terminal, and Text.

The right side of the interface is an R notebook titled "R\_sentiment\_analysis.ipynb". The code in the notebook is:

```
# required packages
packages <- c("SentimentAnalysis", "sentimentr", "corpus", "SnowballC")
#this code only installs the packages listed above if they're not already installed.(which saves time)
install.packages(setdiff(packages, rownames(installed.packages())))

library("domoJupyter")
library("SentimentAnalysis")
library("sentimentr")
library("SnowballC")
library("data.table")
library(tm)
library(dplyr)
library(tidyverse)

library("dplyr")
library("tidyverse")
library("stringr")
library("readr")

library("dplyr")
library("tidyverse")
library("stringr")
library("readr")
```

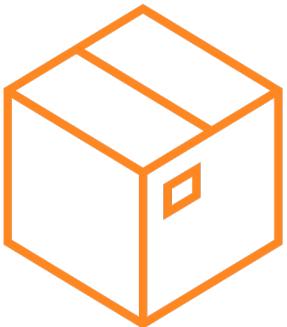
Below the code, there's a section titled "Load in data" with the following code:

```
data <- read_dataframe("Reviews_for_sentiment_analysis", query = "SELECT * FROM `table`", show_col_types = FALSE)
```

A preview of the data is shown as a table:

date	review_id	stars	text	store	city
2022-10-25 11:03:15	ChZDSUhNMG9nS0VJQ0FnSUMtc1B2SUlnEAE	3	(Translated by Google) It's still New York! (Original)	UNIQLO	New York
2022-10-19 04:52:57	ChdDSUhNMG9nS0VJQ0FnSURlczZPvml3RRAB	5	Bought a very nice hat from here.	UNIQLO	New York
2022-10-14 21:13:21	ChdDSUhNMG9nS0VJQ0FnSURlaGVYxzJnRRAB	5	Love UNIQLO. This store has a good selection of clothes. Prices are the same as the store in Brooklyn and N.J.	UNIQLO	New York
2022-10-11 18:29:11	ChdDSUhNMG9nS0VJQ0FnSURlZ2ViehhBRRA	4	A store with huge selection but poor signage. Staff is helpful.	UNIQLO	New York
2022-10-11 10:16:24	ChdDSUhNMG9nS0VJQ0FnSURIM3UtT25BRRA	5	(Translated by Google) Great store, special products (Original) Tolles Geschäft, besondere Produkte	UNIQLO	New York
2022-10-06 08:01:26	ChdDSUhNMG9nS0VJQ0FnSURIN1NnJnRRAB	3	(Translated by Google) It's smaller than the UNIQLO's in Soho or on Fifth Avenue. (Original) Es más pequeña que los UNIQLO del Soho o de la Quinta Avenida	UNIQLO	New York

# Universal Models in Domo



## UNIVERSAL MODELS

- Models that are “one-size-fits-many” & can be used with any data input that fulfills model requirements
- Pre-built & available to all Domo users via one or more of the following: Domo Apps, Cards, Magic ETL Titles
- Business user friendly; no coding required
- Minimal set-up & maintenance
- Good performance, especially as a starting point

## CURRENTLY UNDER DEVELOPMENT

- **Forecasting**
  - Available in September
- **Personally Identifiable Information [PII]**
  - Available in the coming months
- **Sentiment Analysis**
  - Available in the coming months

# Forecasting Universal Model

Vishakha Shenoy

Senior Data Scientist

Team Lead, AI Labs

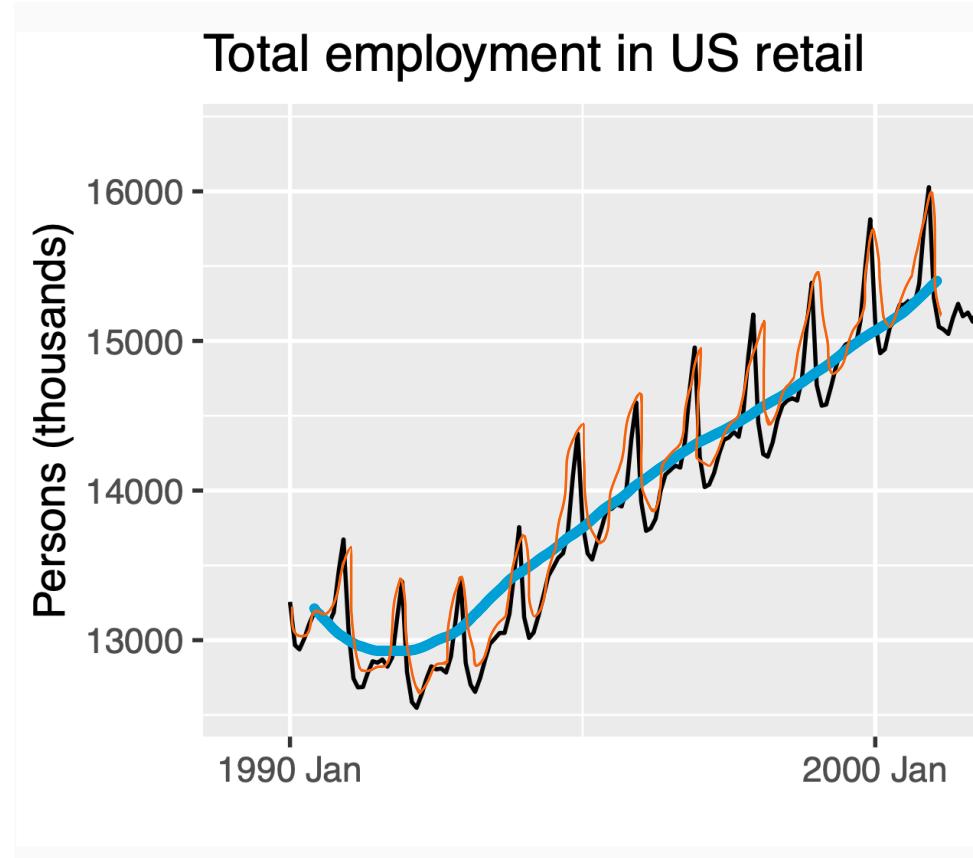
# What is forecasting?

Forecasting involves predicting future outcomes using historical data



# Why use data on the past to predict the future?

Because the future is often like the past



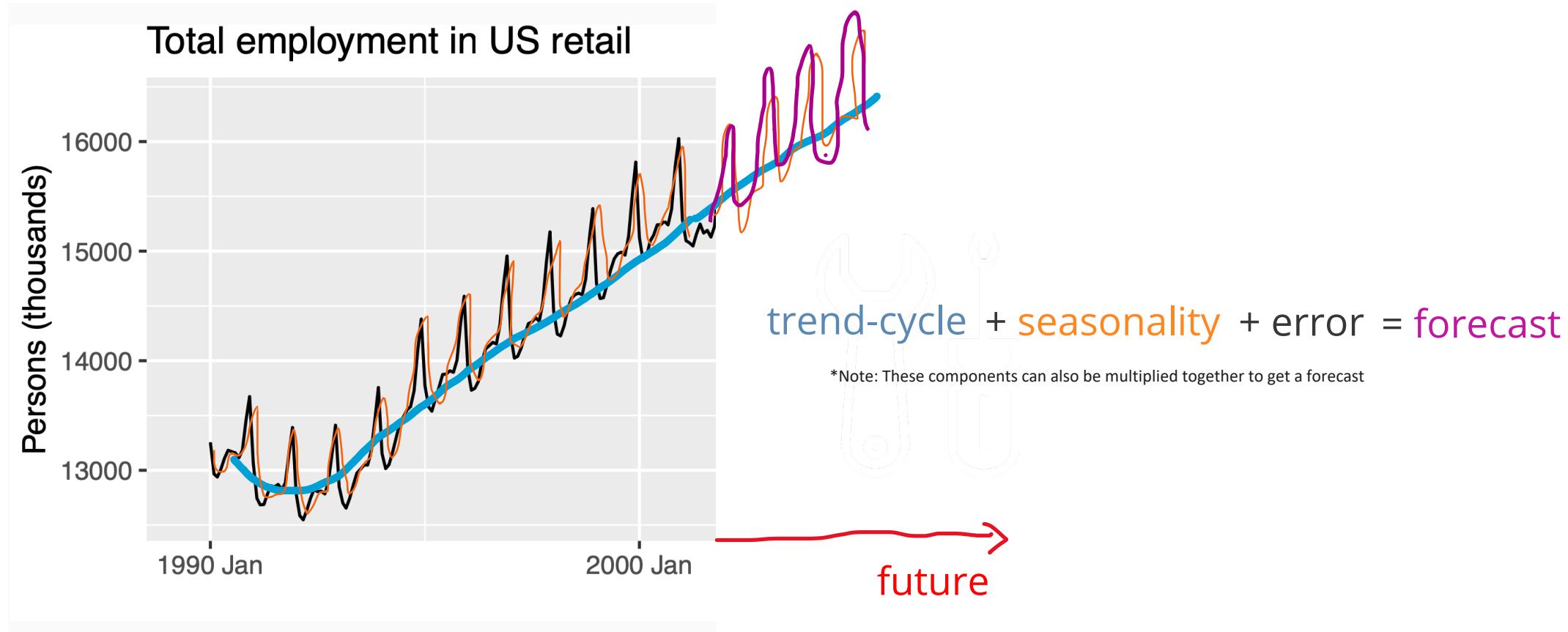
Retail employment tends to follow the same pattern over time

Components of the pattern:

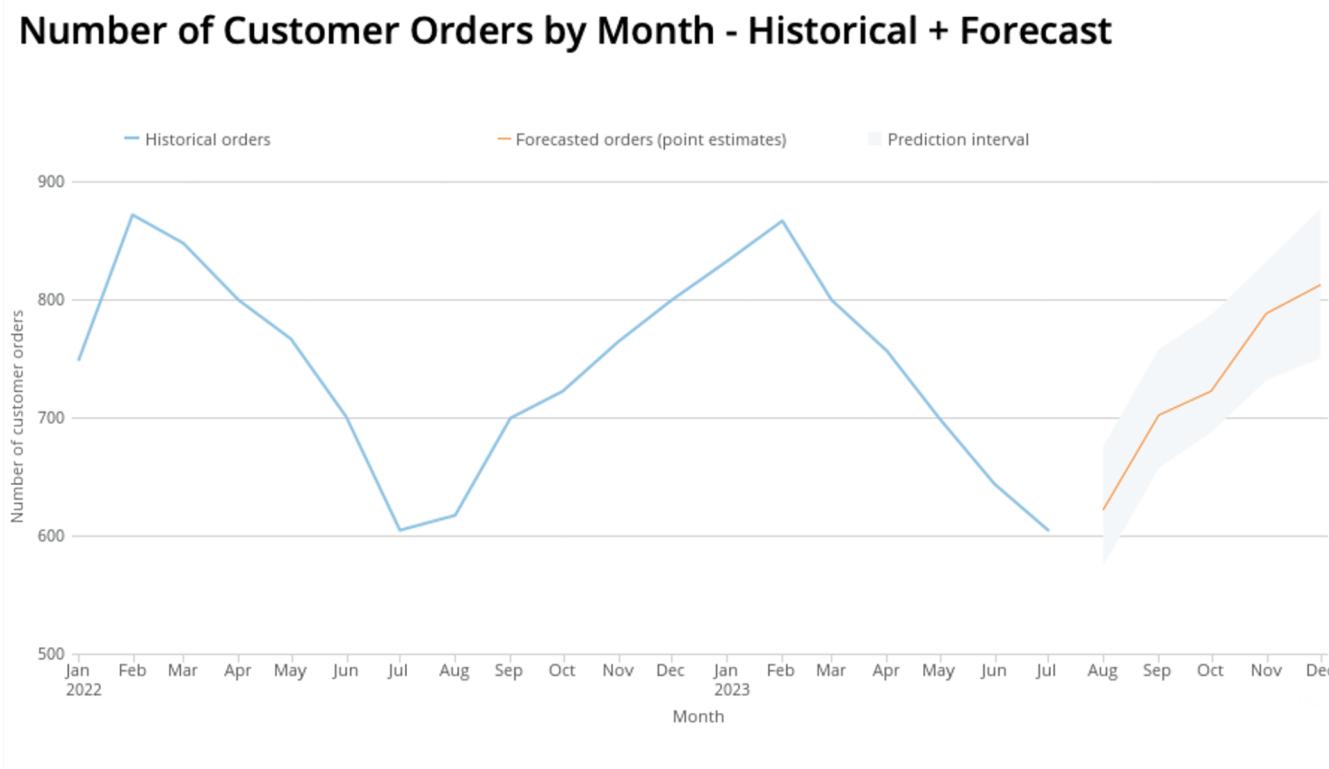
- **Trend-cycle:** long-term increases or decreases in the data
- **Seasonality:** repeating cycle in the series with fixed frequencies (by hour of the day, day of the week, month, etc.)
- **Error/Noise:** leftover fluctuation (variation) in the data

# How Forecasting Works

Forecasting predicts future outcomes by extending historical patterns into the future (using statistics)



# What information do I get from a forecasting model?



- *Prediction intervals*: Range of possible values that future outcome could take with relatively high probability
  - Typical width/range of prediction interval: 95%
- *Point estimates*: Average of possible values that future outcome could take
  - Measured in the same unit as the historical outcome data
- How far in the future you forecast is up to you
  - The farther out you forecast, the more uncertain the future is. Thus, the wider the prediction intervals are

# What kinds of outcomes will Universal Forecast be good at?

Anything that can be observed sequentially over time & has a strong pattern

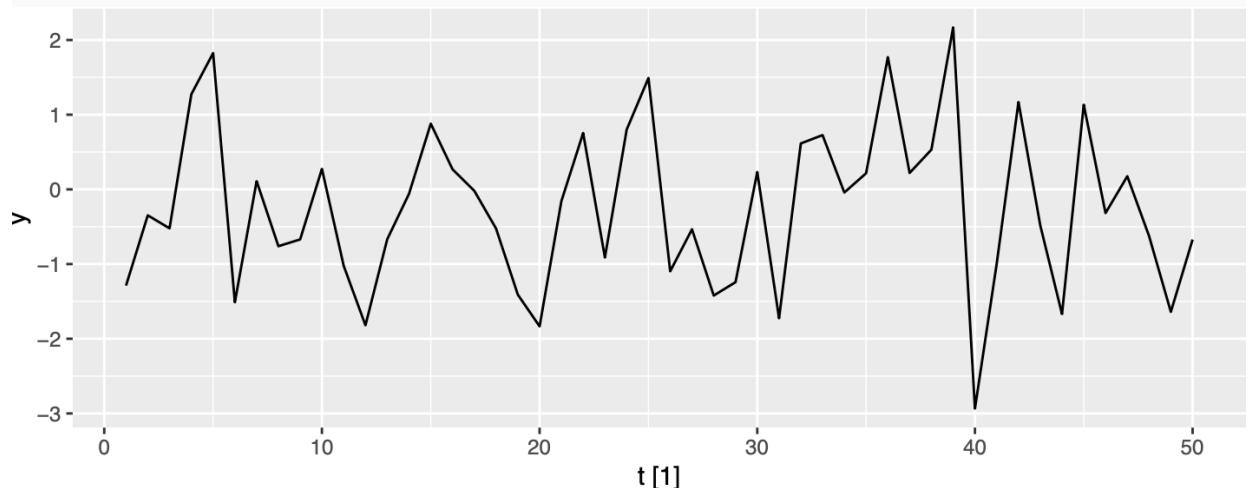
## YES

- Revenue (annual, quarterly, monthly, etc.)
- Sales (annual, quarterly, monthly, etc.)
- Course enrollment (weekly, monthly)
- Resort attendance (daily)

## NO

- Random events (winning lotto numbers, fashion trends, natural disasters, etc.)
- Outcomes with no apparent seasonality or trend

*Outcome with no pattern over time*



# What data do I need in order to forecast?

You need historical data on your outcome that is....

- *Sequential*, meaning observed at regular intervals of time
  - Format of data: 1 column with time variable, 1 column with outcome variable
  - Example time intervals: hourly, daily, weekly, monthly, quarterly, annually
  - The time interval of your forecast = time interval of your historical outcome data
  - Can't have lots of missing data!!

Month	Outcome
24 unique values	123 17 unique values
2022-01-01	749
2022-02-01	871
2022-03-01	848
2022-04-01	799
2022-05-01	767
2022-06-01	701
2022-07-01	605
2022-08-01	617
2022-09-01	700
2022-10-01	722
2022-11-01	764
2022-12-01	800
2023-01-01	832
2023-02-01	866
2023-03-01	800

# What data do I need in order to forecast?

You need historical data on your outcome that is....

- *Plentiful*, meaning enough data to be able to detect patterns (trends + seasonality)
  - What is considered “plentiful” is highly dependent on your outcome, the time interval of observation, and the complexity of the pattern
  - The more data the better!!

Month	Outcome
2022-01-01	749
2022-02-01	871
2022-03-01	848
2022-04-01	799
2022-05-01	767
2022-06-01	701
2022-07-01	605
2022-08-01	617
2022-09-01	700
2022-10-01	722
2022-11-01	764
2022-12-01	800
2023-01-01	832
2023-02-01	866
2023-03-01	800

\*\*This is not enough data to capture yearly trends (only have ~1 year of data) or monthly seasonality (only observe most months once)

# Where would the Universal Forecast be accessible?



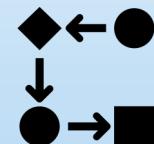
Cards in App Studio



Magic ETL



Jupyter



Workflows

# Universal Forecast Options

Domo's Default  
Model

ARIMA  
ETS

BYOM  
Externally Hosted

# Personal Data Protection Universal Model

Davis Busteed

Senior Data Scientist

AI Labs

# Personal Identifiable Information (PII)

- Any information that can be used to identify a specific individual
- Examples
  - Name
  - Email address
  - Social Security number
  - Street address

# Why PII?

CCPA Privacy Risk  
Regulations GDPR  
PII Security

# Off-The-Shelf PII Services

- Costly
- Difficult to customize
- Optimized for unstructured data

# Domo's Universal Model for PII

## Input data

name abc	department abc	email abc	street_address abc
+ filter		+ filter	+ filter
Dorine Robken	Product Management	drobken0@theatlantic.com	1 Nova Road
Cindee Beswick	Sales	cbeswick1@patch.com	04733 Kedzie Terrace
Ewell Iacavone	Human Resources	eiacavone2@who.int	8 Summit Street
Bartholomeus McKinnon	Research and Development	bmckinnon3@360.cn	2892 Packers Way

# Domo's Universal Model for PII

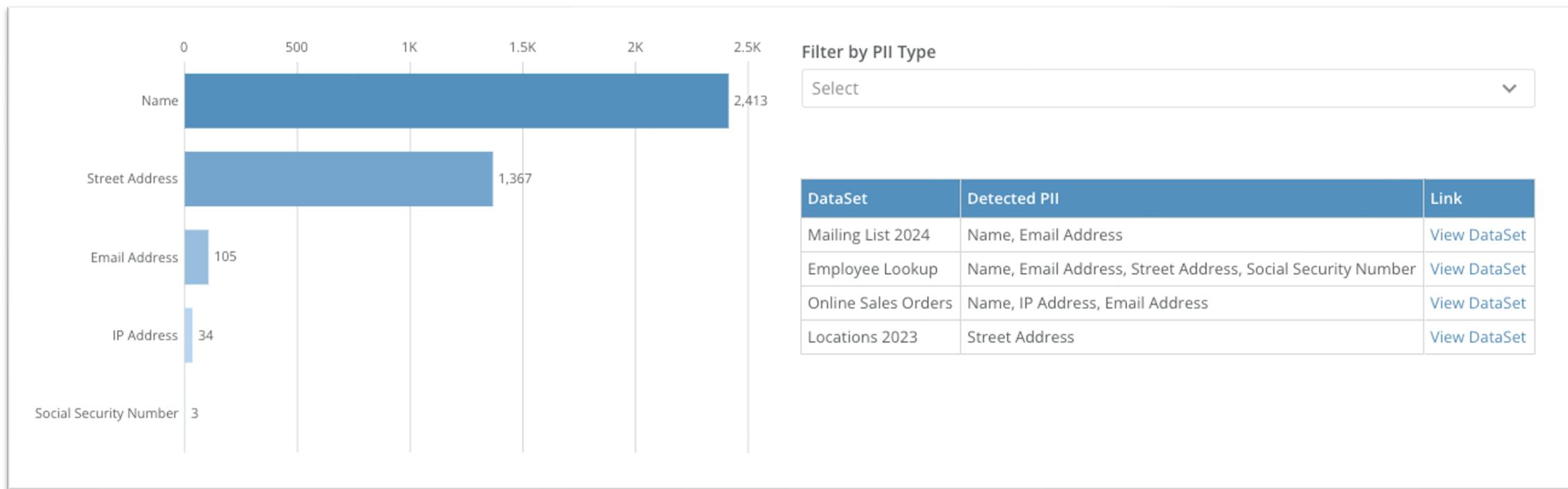
Input data

name	department	email	street_address
Dorine Robken	Product Management	drobken0@theatlantic.com	1 Nova Road
Cindee Beswick	Sales	cbeswick1@patch.com	04733 Kedzie Terrace
Ewell Iacavone	Human Resources	eiacavone2@who.int	8 Summit Street
Bartholomeus McKinnon	Research and Development	bmckinnon3@360.cn	2892 Packers Way

PII results

dataset_id	column	pii_type	confidence_score
1f2bdb7-6dfc-4160-95e8-b86d340ba67a	name	PERSON	0.97
1f2bdb7-6dfc-4160-95e8-b86d340ba67a	department	NONE	
1f2bdb7-6dfc-4160-95e8-b86d340ba67a	email	EMAIL	1.00
1f2bdb7-6dfc-4160-95e8-b86d340ba67a	street_address	ADDRESS	0.94

# PII Reporting

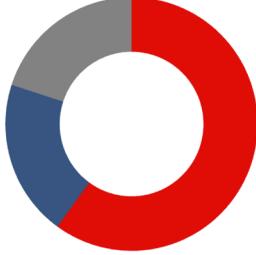


# Universal Model Interface

OVERVIEW DATA CARDS SETTINGS LINEAGE HISTORY PDP ALERTS • AutoML

## Data Profiler

INSIGHTS



DATA OVERVIEW

Total Columns	13	Text	12	92%
Total Rows	1K	Number	1	8%
Total Size	154.93 KB			
Total Datapoints Missing (null)	765 <span>5.9%</span>			

[VIEW STATISTICS BY COLUMN](#) [VIEW FULL SCHEMA](#)

RECOMMENDED ACTIONS

Column	Issue	Action
PII first_name TEXT	Name	<a href="#">HIDE</a> <a href="#">VIEW ISSUE</a>
PII last_name TEXT	Name	<a href="#">HIDE</a> <a href="#">VIEW ISSUE</a>
PII email_address TEXT	Email Address	<a href="#">HIDE</a> <a href="#">VIEW ISSUE</a>

# Other Considerations

- Garbage in, garbage out!

# Sentiment Analysis Universal Model

Bryan Powell

Associate Data Scientist

AI Labs

# Sentiment Analysis



## Generative AI Prompt

You are a product manager tasked with a 3-step process to analyze the following review of a new shoe your clothing company is selling:

<review>

'I like the color of the shoe a lot, it really stands out. The shoe hurts my foot if I walk in it for a long time.'

</review>

<task>

1. Identify up to, but no more than, 3 most important topics in the review.
2. Determine sentiment for each topic in the review: Positive, Neutral, or Negative.
3. Output results as specified within the <outputFormat> XML tags. Do not add any text outside of the specified <outputFormat>. If there are fewer than 3 topics only output distinct topics.

</task>

<outputFormat>

{Topic1: SENTIMENT, Topic2: SENTIMENT, Topic3: SENTIMENT}

</outputFormat>

# Sentiment Analysis

What is it?



- Extracting structured opinions/emotions of an author from unstructured text.

Why is it done?



- Understand your customer base and their needs.
  - Social Media Monitoring
  - Analyzing Reviews
  - Employee Satisfaction Surveys
  - Customer Service Call Transcripts

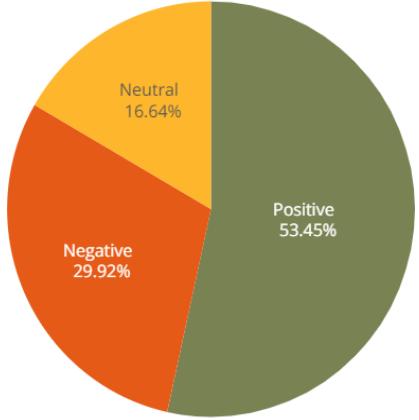
How is it done?



- Manual Analysis
  - Small samples
- Machine learning models
  - Universal models
    - One size fits all.
  - Custom models
    - Several approaches
    - Several levels of depth

## Overall Sentiment

*Reviews are most often positive*



## Overall Rating

*The average review rating is 3.15*



## Number of Reviews

1.19K

## Live feed of recent reviews

Sentiment	Rating	Review	Location
Positive	★★★★★	the manager was really nice and accommodating. the store is well kept and super organized. love shopping during the annual sales you can get great deals!	Portland, OR
Positive	★★★★★	great company. i found a ton of great items during their after holiday sale. well organized and clean compared to other locations	Miami, FL
Neutral	★★★★★	they have a lot of new arrivals	Springfield, IL
Negative	★	the worst, most rude customer service of any corporation i've ever been to. shame on you zahra! every location is awful and this one is the worst	Portland, OR
Positive	★★★★★	very nice store of the brand	Burbank, CA
Negative	★	the worst store ever. lines everywhere.	Portland, OR
Positive	★★★★★	all very beautiful very comfortable	American Fork, UT

## Basic Sentiment

- **Sentiment Model assigns a polarity score to each review.**
- **Assignment of labels can be adjusted for accuracy.**

## Sentiment by topic

Prices

160 # of reviews



Positive	65%
Negative	18%
Neutral	18%

Selection

295 # of reviews



Positive	79%
Negative	11%
Neutral	10%

Staff

343 # of reviews



Negative	54%
Positive	39%
Neutral	7%

## Live feed of recent reviews

Sentiment	Rating	Review	Location
Positive	★★★★★	the manager was really nice and accommodating. the store is well kept and super organized. love shopping during the annual sales you can get great deals!	Portland, OR
Negative	★★	i'm your loyal customer . i love your clothes. but today.. customer service leas the worst one. young kids should know how to deal with business. or you should keep eyes on young employees.... no more.. sorry.	Denver, CO
Negative	★	customer service at this company is horrible, they don't care about the physical condition of their customers. i'm on crutches with my injured foot in a huge line waiting (i asked an employee if i would have the preferential line he said no) my foot is swollen and hurting. two days ago i was in miami shopping at company, with crutches and i was very well attended and they gave me preference when i went to pay. last time i enter this company. thanks for not care your customers. terrible	Miami, FL
Positive	★★	this company always has very long lines to buy things. the staff are friendly but i wonder	Springfield, IL

## Sentiment by Topic

- Identify topics of interest.
- Apply classification model to reviews.
  - Pulled model from Hugging Face.
- Identify drivers of positive/negative reviews

## Sentiment Using LLMs

- Takes customization to another level.
- Prompt Engineering is key.
  - Use XML tags to parse prompt.
  - Provide task in step-by-step format.
  - Specify output format.
  - Provide LLM with examples.

### Generative AI Prompt (Domo AI Playground)

You are a product manager tasked with a 3-step process to analyze the following review of a new shoe your clothing company is selling:

```
<review>  
'I like the color of the shoe a lot, it really stands out. The shoe hurts my foot if I walk in it for a long time.'  

```

```
<task>  
1. Identify up to, but no more than, 3 most important topics in the review.  
2. Determine sentiment for each topic in the review: Positive, Neutral, or Negative.  
3. Output results as specified within the <outputFormat> XML tags. Do not add any text outside of the  
specified <outputFormat>. If there are fewer than 3 topics only output distinct topics.  

```

```
<outputFormat>  
{Topic1: SENTIMENT, Topic2: SENTIMENT, Topic3: SENTIMENT}  

```

### Generative AI Prompt (Domo AI Playground)

```
4 v {  
5   "output": "<outputFormat>\n{Color: Positive, Comfort: Negative}\n</outputFormat>"  
6 }
```