Forecasting Using Domo

Lunch & Learn | April 6, 2023

Florencia Silveira, Bryan Powell, & Steven Guitron

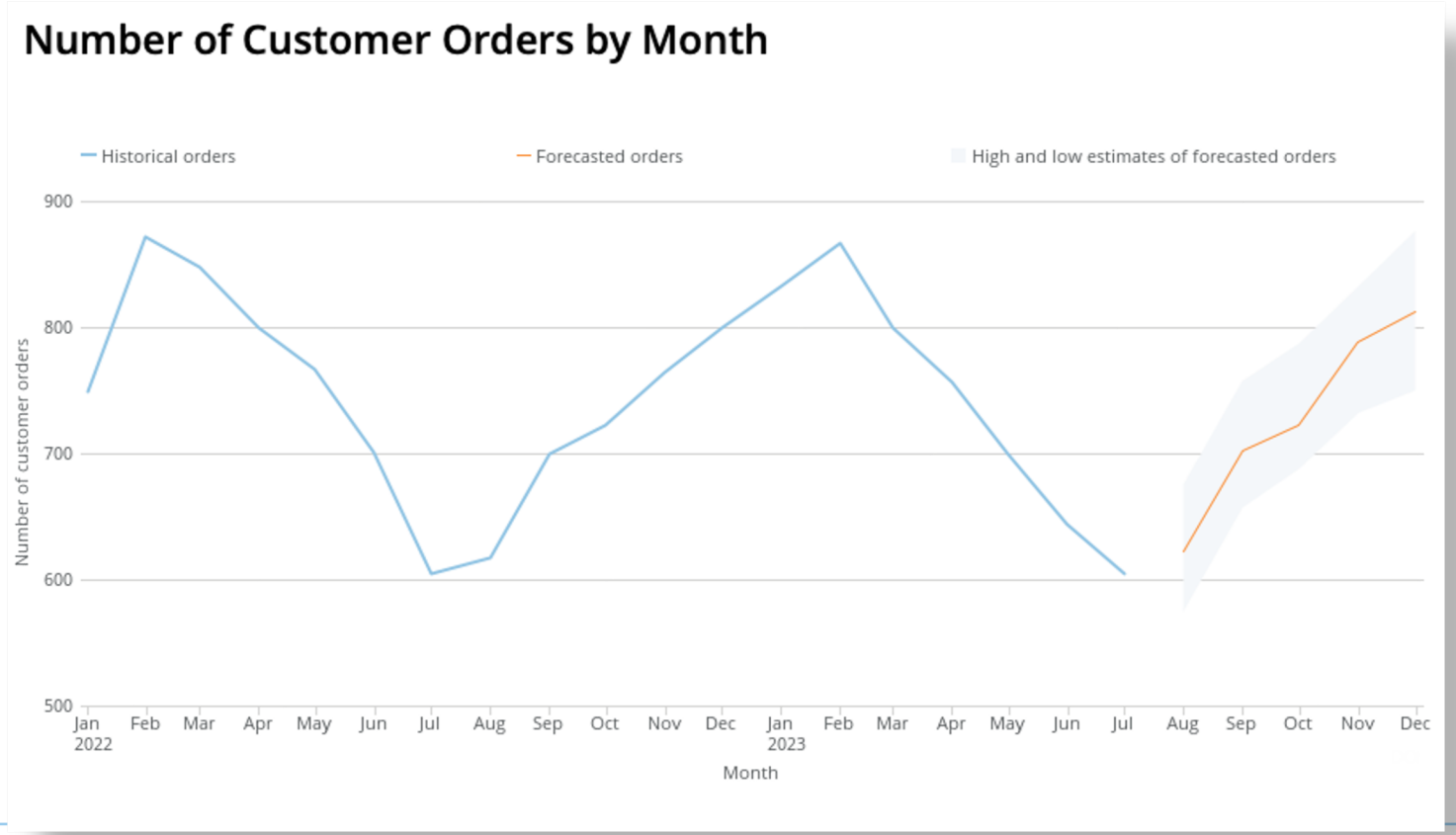
Data Scientists

Data Science Professional Services at Domo, Inc.

What is forecasting?

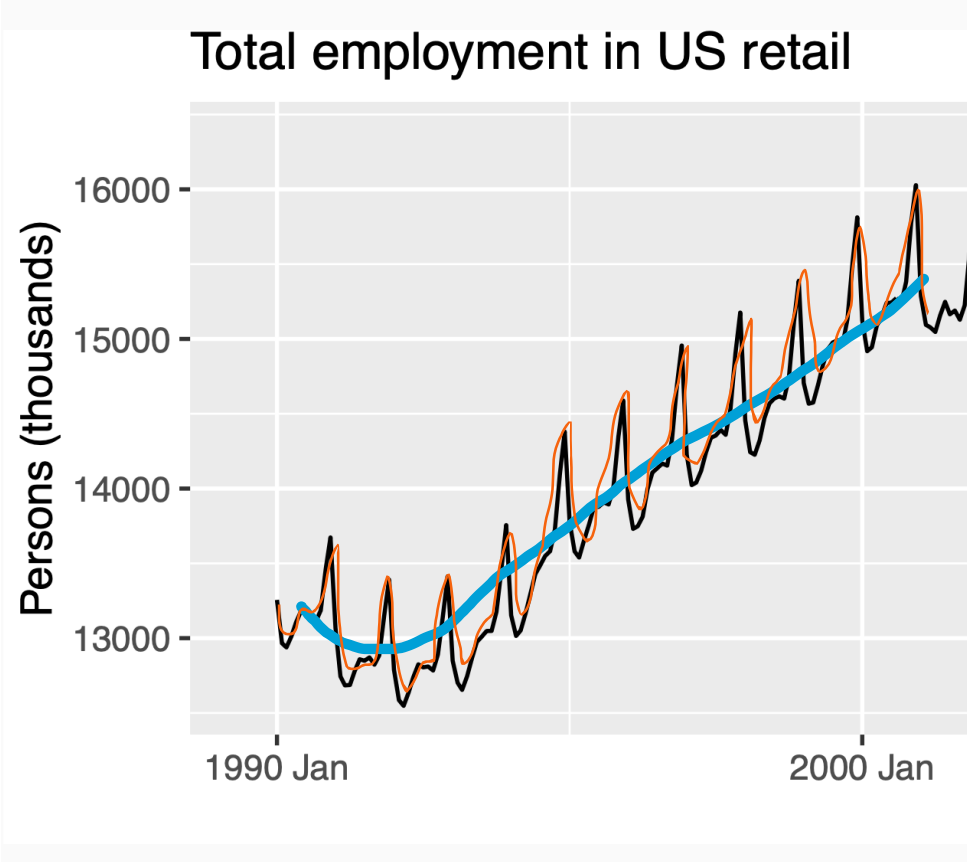
Forecasting involves predicting future outcomes using historical data

- Specific focus of this presentation: forecasting using historical *outcome* 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

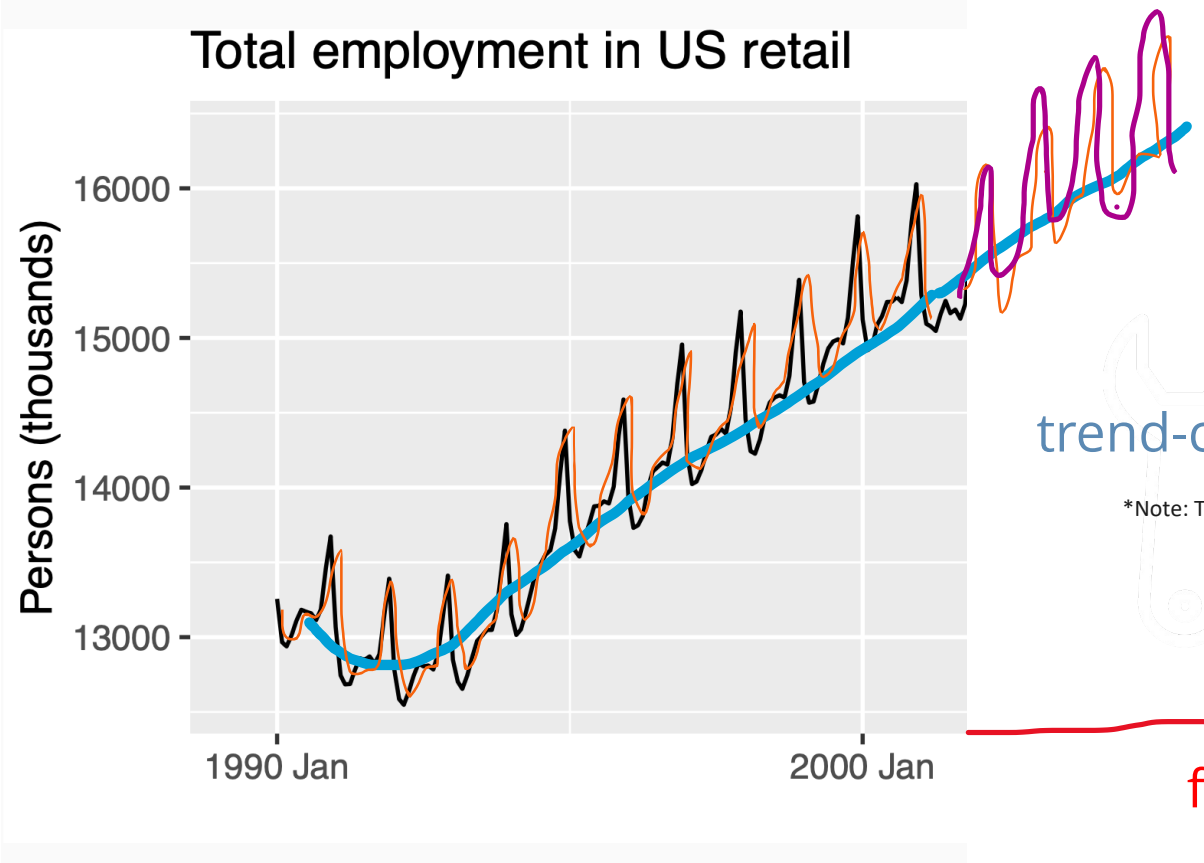
Components of the pattern:

- 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

Figure used with permission from: Hyndman, R.J., & Athanasopoulos, G. (2021) Forecasting: principles and practice, 3rd edition, OTexts: Melbourne, Australia. OTexts.com/fpp3.

How Forecasting Works

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

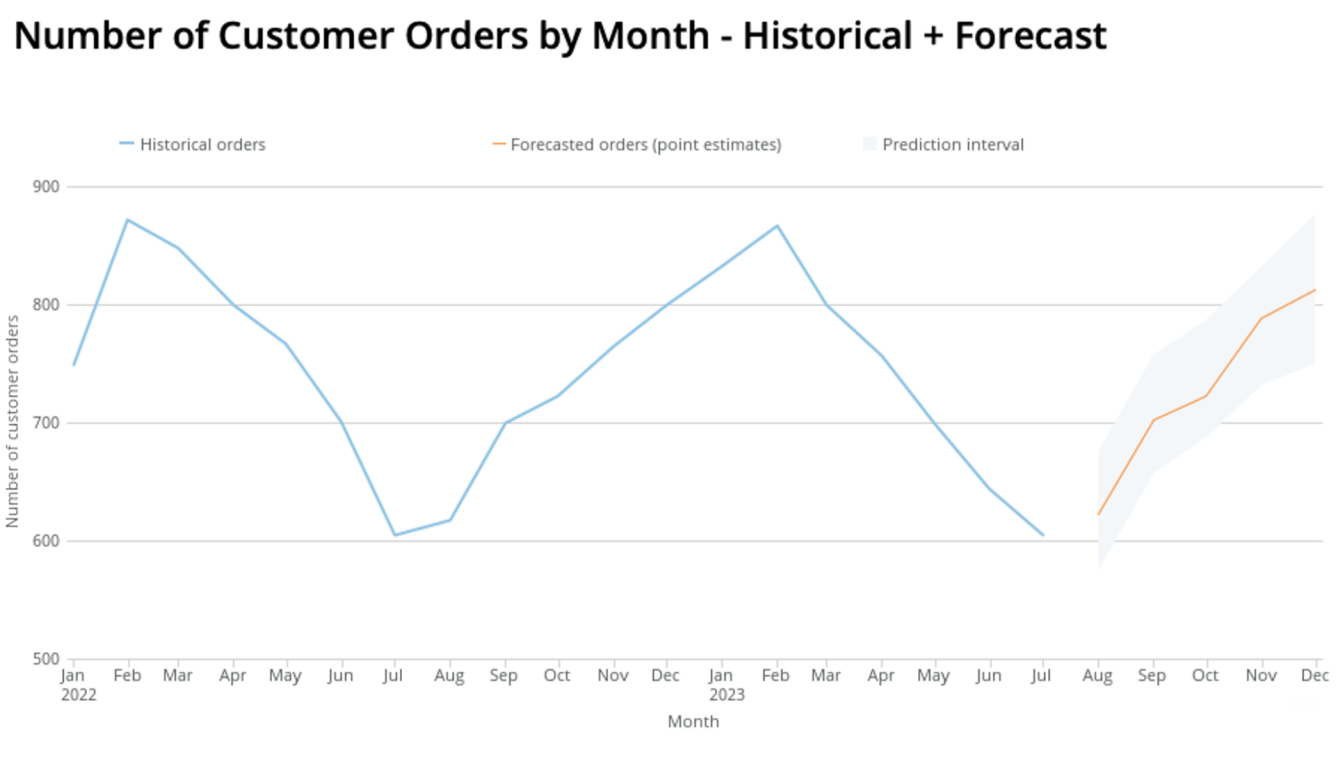


trend-cycle + seasonality + error = forecast

*Note: These components can also be multiplied together to get a forecast

Figure used with permission from: Hyndman, R.J., & Athanasopoulos, G. (2021) Forecasting: principles and practice, 3rd edition, OTexts: Melbourne, Australia. OTexts.com/fpp3.

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 can be forecast?

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

YES

- Annual company revenue
- Quarterly customer purchase amounts
- Weekly course enrollment
- Daily resort attendance

NO

- Random events (winning lotto numbers, fashion trends, natural disasters, etc.)

Outcome with no pattern over time

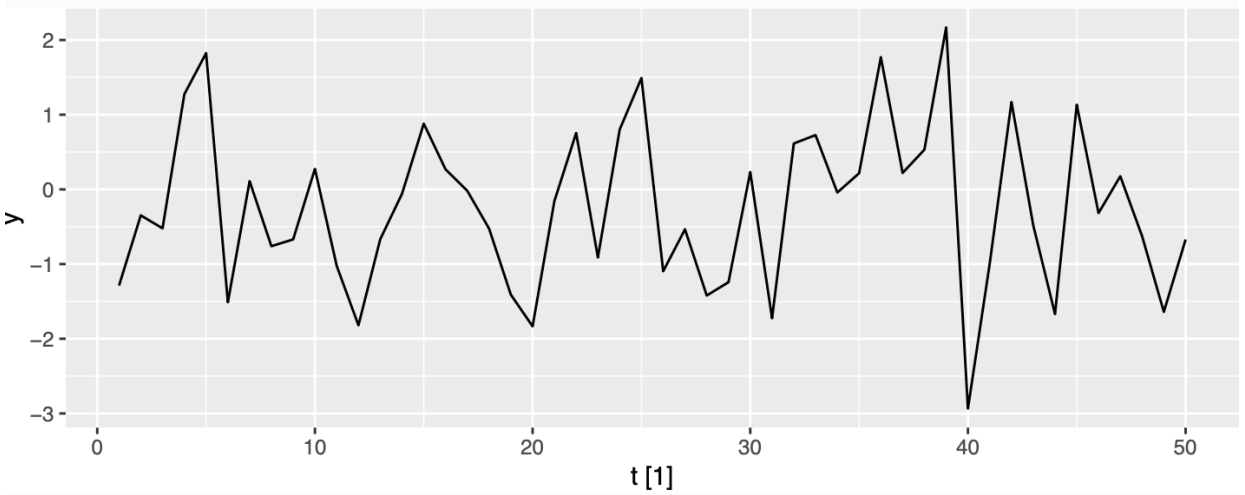


Figure used with permission from: Hyndman, R.J., & Athanasopoulos, G. (2021) Forecasting: principles and practice, 3rd edition, OTexts: Melbourne, Australia. OTexts.com/fpp3.

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 24 unique values	Outcome 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 24 unique values	Outcome 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

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

Forecasting using Domo

	Forecasting Tools	Availability	Ease of Use	Forecasting Output		Customization		Forecasting Visualization Options
				Point estimates	Prediction intervals	Available algorithms	Assess accuracy using train/test samples	
Bronze	***Line and bar chart cards: last value projection or multi-period projection option*** More info	Available to all customers	Very easy	Yes	No	Linear regression, forecasting, averages (you specify)	No	Forecast displayed as part of line or bar chart cards
Silver	***Forecasting Magic ETL Data Science tile*** More info	Premium feature, available as part of Data Science toolset	Moderate	Yes	Yes	ARIMA	No	Forecasting chart card
Silver	***AutoML*** More info	Premium feature, available as part of Data Science toolset	Moderate	Yes	No	Linear Learner, XGBoost, multilayer perceptron (MLP) (the tool chooses)	Yes (automatically provided, but not customizable)	Forecasting chart card
Gold	***Jupyter Workspaces*** More info	Premium feature, available as part of Data Science toolset	More challenging (requires knowledge of how to code in R or Python)	Yes	Yes	Any forecasting algorithm you specify	Yes (fully customizable)	Forecasting chart card OR custom visualization created in Jupyter

Live Demos of Domo Forecasting Tools

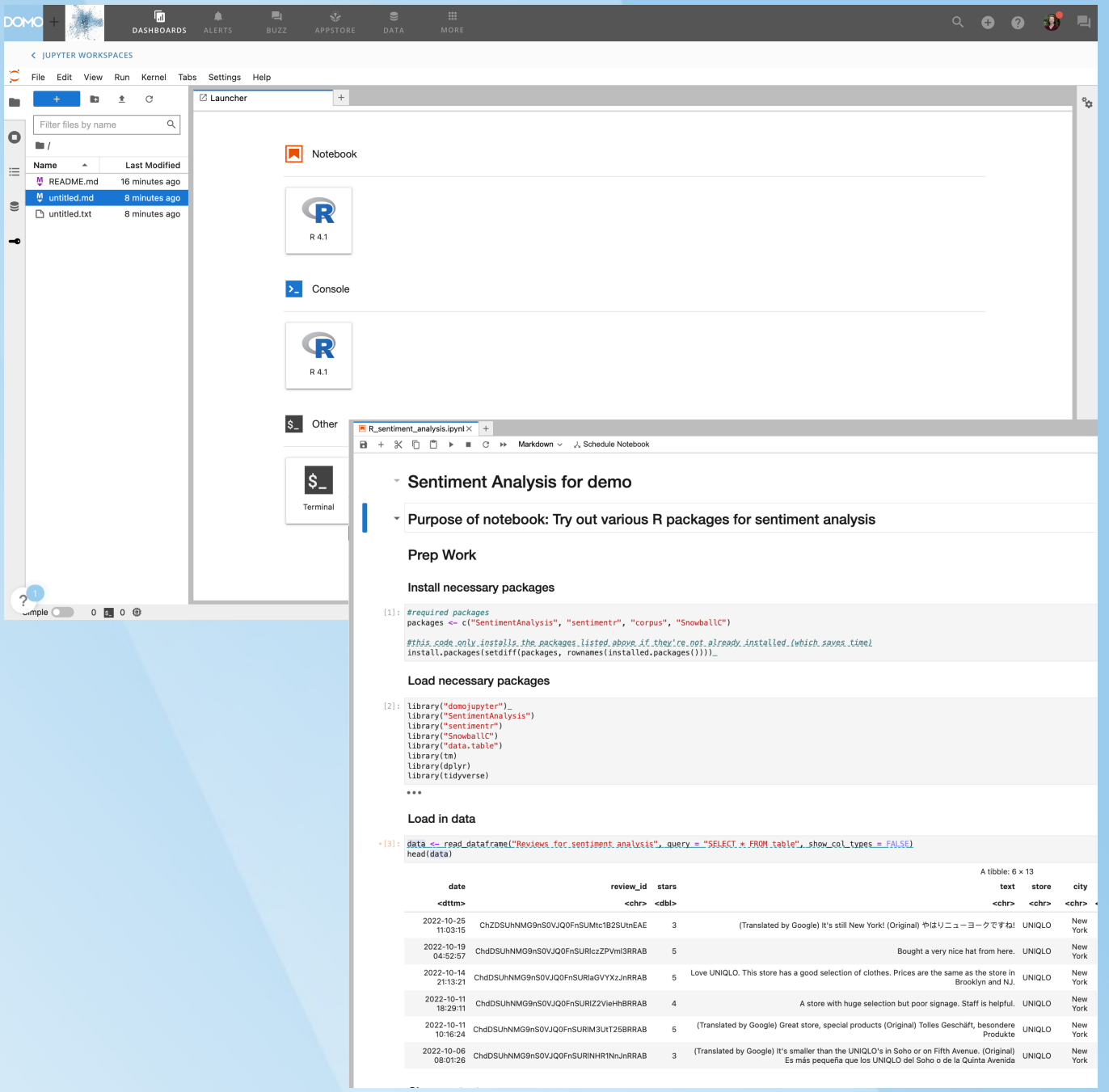
Thank You

Jupyter Workspaces

Domo's integrated Jupyter Workspaces let's you explore your data, train & productionize models, and create custom visualizations - all in one place

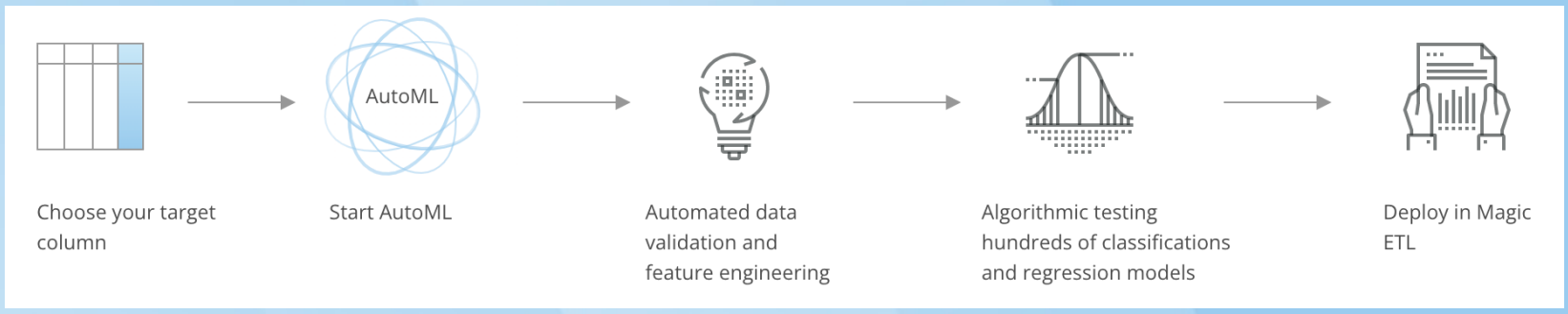
FEATURES

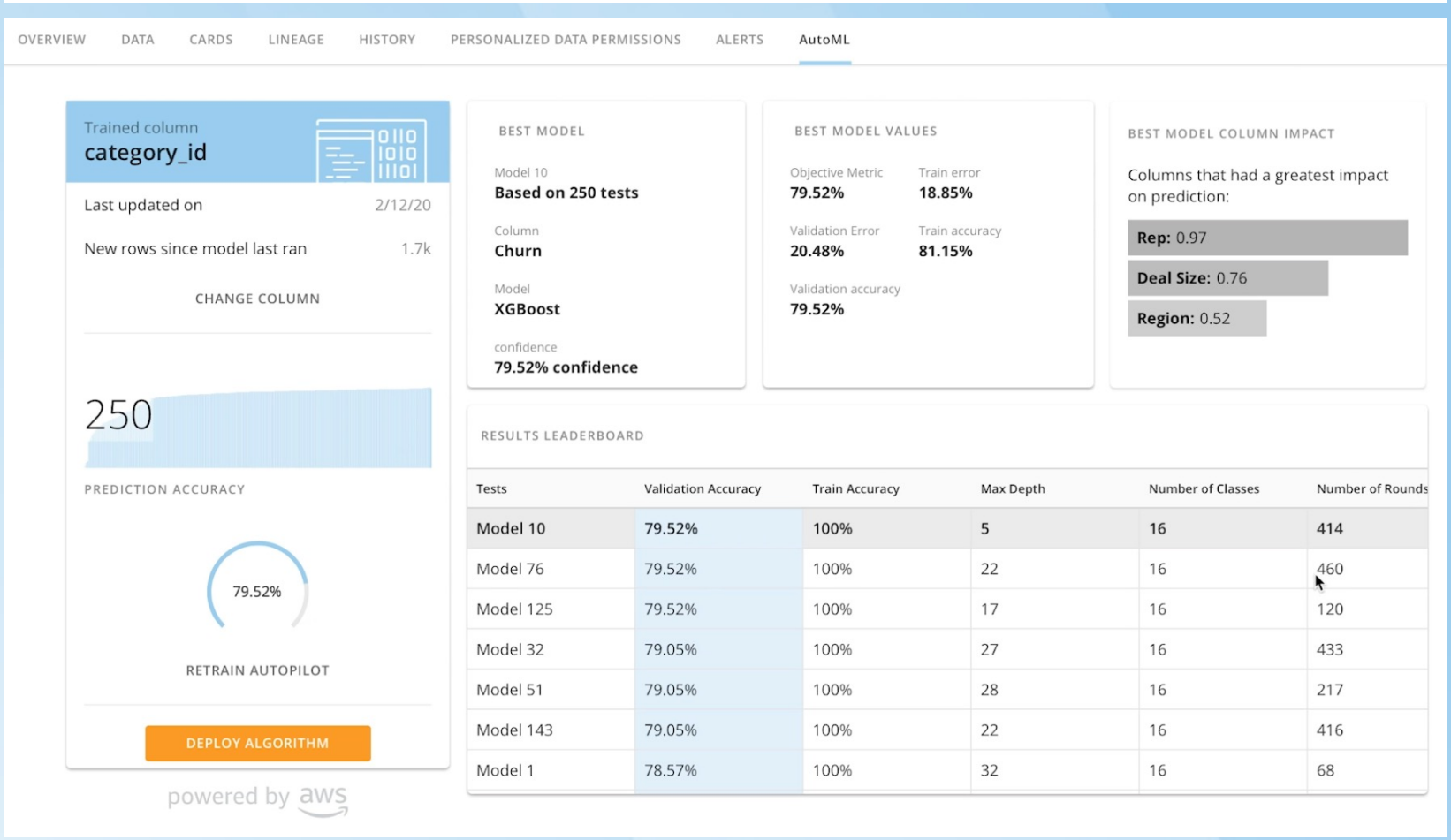
- Read & write data directly to & from Domo
- Write code in notebooks using Python or R programming languages
- Choose your compute tier
- Productionize models in a notebook
- Run tasks programmatically (for developers)
- Schedule notebooks to run automatically at a set time or frequency
- Share workspaces to enable collaboration
- Integrate with GitHub for version control



AutoML

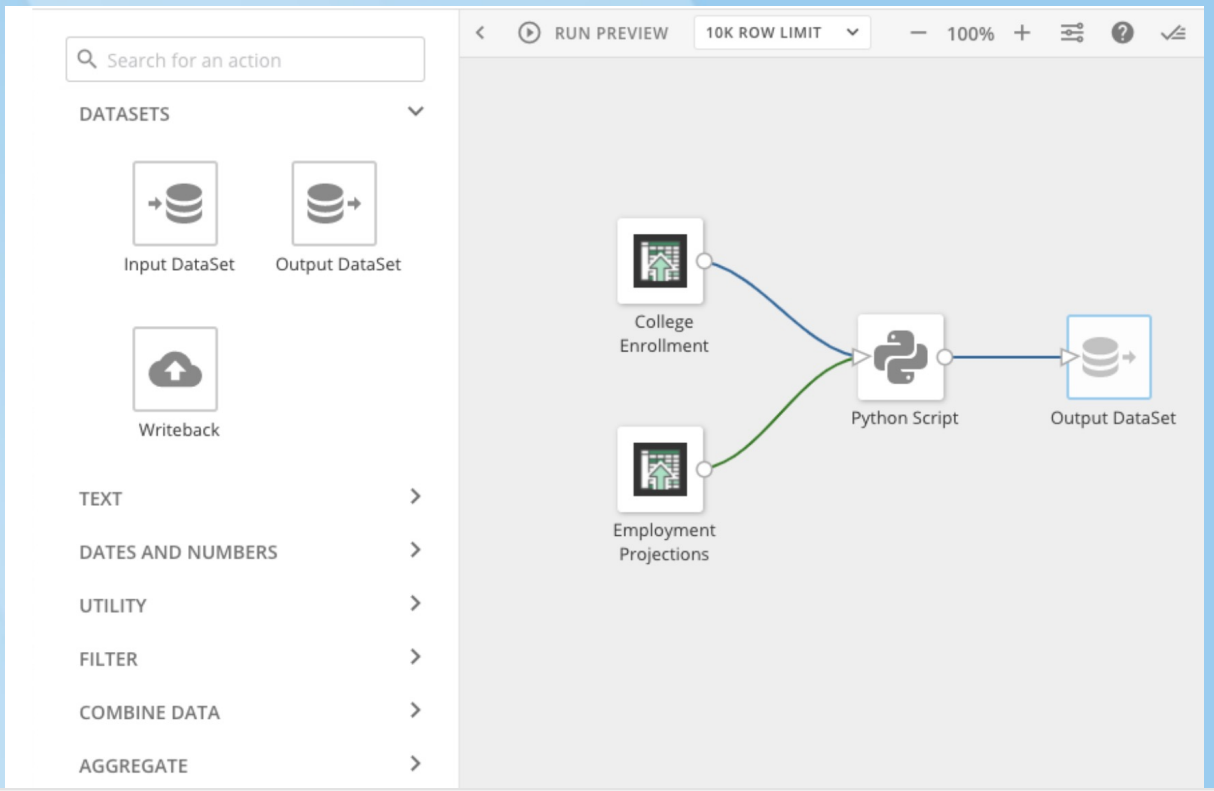
AutoML is a low-code option for automatically generating & testing multiple machine learning models





Scripting Tiles

Use Scripting Tiles to run brief R or Python code directly in an ETL dataflow



Python Script

CODE SCHEMA NOTES

```
1  # Import the domomagic package into the script
2  from domomagic import *
3
4  # read data from inputs into a data frame
5  input1 = read_dataframe('College Enrollment')
6
7  # write your script here
8
9
10 # write a data frame so it's available to the next action
11 write_dataframe(input1)
```