Analytics mindset

Alternative data

Apollo Investment Research

Note: It is recommended that you complete the EYARC Analytics mindset case study, Alternative data overview, before completing this case.

Overview

This case focuses on financial statement forecasting and analysis. In completing this case, you will learn more about how companies can use alternative data — such as satellite imagery of retail parking lots — to improve the speed and accuracy of their decision-making.

Background

You are a financial analyst for Apollo Investment Research (Apollo), a company that develops research reports and stock recommendations for midsize institutional investor clients, including pension funds, mutual funds and university endowments. Apollo strives to use cutting-edge technology and data to provide the most profitable recommendations possible to its clients. In order to accomplish this, the company regularly assesses new data sets and technologies to determine whether they materially improve financial forecasts, valuations and investment recommendations.

Over the past 10 years, you have followed the retail industry, producing investment research for clients about Target, Walmart, Home Depot and Lowe's. Your investment research on these companies typically includes: (1) forecasts of key performance metrics, such as sales and earnings; (2) valuations of the companies; and (3) investment recommendations, such as buy, hold or sell.

For these retail chains, a key performance metric is same-store sales growth (also referred to as comparable store sales growth). This metric focuses on store locations that have been operating for more than a year and captures how much sales have grown in those stores during the period (typically, a quarter) compared to the same period in the prior year. For example, in the first quarter of 2022, Target reported 3.3% same-store sales growth, meaning that sales in the second quarter of 2022 have increased by 3.3% relative to sales in the first quarter of 2021 for Target store locations that have been operating for more than a year.

Same-store sales growth is the most widely reported and followed metric in the retail industry because it plays a crucial role in forecasting and valuation. This is the metric you should focus on first in your forecasting and valuation models because it is key to the accuracy of forecasts and valuations and the profitability of investment recommendations.

SCORE no. 16691-221US_4

Part 1: Ask the right questions

Recently, your boss Maria asked you to evaluate a new data set that was recently purchased for $20,000 by Apollo to determine whether it can improve your investment research and recommendations. This data set was produced by a company that launched a collection of nanosatellites into space. The nanosatellites daily produce high-resolution images of thousands of store parking lots for several retail chains (see Figure 1 for an example[1]), including the ones you follow — Target, Walmart, Home Depot and Lowe's. Automated counting software calculates the number of cars in each parking lot, as well as the number of spaces in each parking lot.

Figure 1: Nanosatellite image of retail parking lot.

The nanosatellite data set includes seven columns of information for eight years (2011–18) for each company (between approximately 240,000 and 340,000 lines of data each), including:

1. Ticker: company stock ticker (e.g., Walmart is WMT)
2. Type: store ID for the specific store location captured in the image
3. State: abbreviation of the state for the store location captured in the image
4. ZIP: ZIP code for the store location captured in the image
5. Notes: the date the image was captured
6. Cars: the number of cars in the parking lot image
7. Spaces: the number of spaces in the parking lot image

[1] Image used with permission from RS Metrics. "TrafficSignals™ Historical Dataset | Retail Monitoring," *RS Metrics*, https://learn.rsmetrics.com/trafficsignals/retail/monitoring/special/offer?hsLang=en, accessed June 2022.

SCORE no. 16691-221US_4

Your boss Maria has asked that you evaluate whether the data set can materially improve your financial forecasts, valuations and investment recommendations.

Required

- Brainstorm different ways you could use this data set to improve your financial forecasts, valuations and investment recommendations.
  - For each idea, identify any limitations in the nanosatellite data that might prevent the data from being useful in the way you expect.
  - For each idea, explain how you could use the data provided by Maria to test whether or not the data set improves your financial forecasts, valuations and investment recommendations. Additionally, identify any other data that would be needed to carry out your tests.
- Can you think of any other data sources or technology that could accomplish the same thing as the nanosatellite data set, but might be potentially cheaper or more effective? Explain your answer and any additional considerations.

SCORE no. 16691-221US_4

Analytics mindset

Alternative data

Apollo Investment Research

Part 2: Extract, transform and load relevant data

In order to evaluate the usefulness of the nanosatellite data set, you have decided to try to answer three key questions:[2]

- Does the nanosatellite data improve my forecasts of same-store sales growth above and beyond other predictors?
- Does the nanosatellite data predict stock returns around earnings announcements (when companies announce financial performance to the public) above and beyond other predictors?
- Is the nanosatellite data more or less useful during different seasons of the year (e.g., is the nanosatellite data more or less predictive of same-store sales growth during the busy holiday season for retailers, including November, December and January, or during the non-holiday season)?

You have decided that answering these three questions will give you some insight into the broader question of whether the nanosatellite data can materially improve your financial forecasts, valuations and investment recommendations.

In order to answer these questions, you have been provided a spreadsheet from a coworker that combines data from three sources. Each observation in the spreadsheet is a company-quarter observation, meaning it captures quarterly data for each company from 2012 through 2018. The data sources and key variables are listed below.

- Nanosatellite database:
  - **Unique ID:** a unique identifier for each line of data in the data set
  - **Ticker:** each company's stock ticker
  - **Cars**: the number of cars that occupy spaces aggregated across all days and store locations during the quarter
  - **Spaces:** the number of available spaces aggregated across all days and store locations during the quarter
  - **LagCars**: the number of cars that occupy spaces aggregated across all days and store locations during the same quarter for the year prior
  - **LagSpaces:** the number of available spaces aggregated across all days and store locations during the same quarter for the year prior

[2] Some of the key questions in this case were inspired by academic research on nanosatellite imagery located at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3222741.

SCORE no. 16691-221US_4

- Company earnings announcements:
  - **Date:** the last day of the fiscal quarter covered by the earnings announcement
  - **DeltaCompSales:** growth in same-store sales growth for the quarter relative to the same quarter for the year prior
  - **LagSales:** the prior year's same-store sales growth for the quarter relative to the same quarter two years prior
- Market data:
  - **Qreturn:** buy-and-hold stock return aggregated over the quarter
  - **EAreturn:** three-day buy-and-hold stock return surrounding the earnings announcement for the quarter
  - **Holiday:** variable takes the value of 1 for quarters that take place during retailers' busy holiday months (November, December and January) and 0 otherwise

Required

- Obtain the data file **Analytics_mindset_case_studies_Alternative_data_ApolloP2.xlsx**.
  - Create a new field **ParkingRate** and **LagParkingRate** that calculates the percentage of cars to spaces for the quarter (**Cars**/**Spaces**) and for the same quarter for the year prior (**LagCars**/**LagSpaces**). This is known as the parking lot fill rate.
    - For context, the average industry parking lot fill rate is 28.34%, meaning that parking lots are just over one-quarter full on average.
  - Create a new field **DeltaParkingRate** that captures the growth in the company's parking lot fill rate relative to the same quarter for the year prior ((**ParkingRate**-**LagParkingRate**)/**LagParkingRate**).
    - For context, the average growth in parking lot fill rate is 1.4% during the period covered in this data set.
  - Transform and load the data into the appropriate software as directed by your instructor.
- Before performing any analyses on your data, develop an expectation of what you anticipate the analyses will reveal for each key question. Explain your reasoning.

SCORE no. 16691-221US_4

Analytics mindset

Alternative data

Apollo Investment Research

Part 3.1: Apply appropriate data analytics techniques

Required

Perform the necessary analyses to answer the first key question: Does the nanosatellite data improve my forecasts of same-store sales growth above and beyond other predictors?

- Using your visualization software as directed by your instructor (e.g., Power BI, Tableau, Spotfire), prepare an analysis of the relationship between growth in the parking lot fill rate and growth in same-store sales by creating a scatterplot of the two variables.
  - Each company-quarter observation should be represented as a point on the scatterplot.
  - Include a trend line. Write the equation for the trend line and provide the R^2.
  - Create an additional scatterplot that compares this relationship across the four different companies you follow.
- Use a software program as directed by your instructor (e.g., Alteryx, Excel, SAS, R) to conduct a linear regression analysis that predicts the growth in same-store sales as a function of growth in the parking lot fill rates.
  - Control for other important predictors of same-store sales growth, including:
    - Control for prior year same-store sales growth. (Note: This is a known predictor of same-store sales growth.)
    - Control for stock returns during the quarter. (Note: This controls for other signals that investors incorporate into stock prices, such as other important company announcements or events that occurred during the quarter.)

SCORE no. 16691-221US_4

Analytics mindset

Alternative data

Apollo Investment Research

Part 3.2: Apply appropriate data analytics techniques

Required

Perform the necessary analyses to answer the second key question: Does the nanosatellite data predict stock returns around earnings announcements (when companies announce financial performance to the public) above and beyond other predictors?

- Using your visualization software as directed by your instructor (e.g., Power BI, Tableau, Spotfire), conduct a visual analysis of the relationship between growth in the parking lot fill rate and stock returns around earnings announcements.
  - Each company-quarter observation should be represented as a point on the scatterplot.
  - Include a trend line. Write the equation for the trend line and provide the R^2.
  - Create an additional scatterplot that compares this relationship across the four different companies you follow.
- Use a software program as directed by your instructor (e.g., Alteryx, Excel, SAS, R) to conduct a linear regression analysis that predicts returns around earnings announcements as a function of growth in the parking lot fill rates.
  - Control for other important predictors of earnings announcement returns, including:
    - Control for prior year same-store sales growth. (Note: This is a known predictor of same-store sales growth.)
    - Control for stock returns during the quarter. (Note: This controls for other signals that investors incorporate into stock prices, such as other important company announcements or events that occurred during the quarter.)

SCORE no. 16691-221US_4

Analytics mindset

Alternative data

Apollo Investment Research

Part 3.3: Apply appropriate data analytics techniques

Required

Perform the necessary analyses to answer the third key question: Is the nanosatellite data more or less useful during different seasons of the year (e.g., is the nanosatellite data more or less predictive of same-store sales growth during the busy holiday season for retailers, including November, December and January, or during the non-holiday season)?

- Using your visualization software as directed by your instructor (e.g., Power BI, Tableau, Spotfire), conduct a visual analysis of the relationship between growth in the parking lot fill rate and growth in same-store sales by creating a scatterplot of the two variables. The scatterplot should compare this relationship across holiday quarters and non-holiday quarters.
  - Each company-quarter observation should be represented as a point on the scatterplot.
  - Include a trend line.
- Use a software program (e.g., Alteryx, Excel, SAS, R) to conduct a linear regression analysis that predicts growth in same-store sales as a function of growth in the parking lot fill rates across holiday vs. non-holiday quarters. You can accomplish this by analyzing the interaction effects. In regression, an interaction effect indicates that the effect of an independent variable on a dependent variable changes, depending on the values of another independent variable. For this question, you want to examine whether the effect of DeltaParkingRate on DeltaCompSales changes, depending on the values of Holiday. As such, you should focus on the interaction variable Holiday*DeltaParkingRate.
  - The regression should include the following independent variables of interest: Holiday, DeltaParkingRate and the interaction of the two (Holiday*DeltaParkingRate).
    - If using Alteryx, create a new variable, Holiday*DeltaParkingRate, that is the product of the Holiday and DeltaParkingRate variables. Ensure this variable is encoded as a numerical variable. Include this new variable in the regression, along with Holiday and DeltaParkingRate.
  - Control for other important predictors of same-store sales growth, including:
    - Control for prior year same-store sales growth. (Note: This is a known predictor of same-store sales growth.)
    - Control for stock returns during the quarter. (Note: This controls for other signals that investors incorporate into stock prices, such as other important company announcements or events that occurred during the quarter.)

SCORE no. 16691-221US_4

Analytics mindset

Alternative data

Apollo Investment Research

Part 4: Interpret and share the results with stakeholders

Required

- Interpret the results of the visual analysis and regression for the first key question: Does the nanosatellite data improve my forecasts of same-store sales growth above and beyond other predictors?
  - Interpret the regression coefficient on the DeltaParkingRate variable.
  - Interpret the P value on the DeltaParkingRate variable.
  - Did the results of your analysis align with your expectations?
- Interpret the results of the visual analysis and regression for the second key question: Does the nanosatellite data predict stock returns around earnings announcements (when companies announce financial performance to the public) above and beyond other predictors?
  - Interpret the regression coefficient on the DeltaParkingRate variable.
  - Interpret the P value on the DeltaParkingRate variable.
  - Did the results of your analysis align with your expectations?
- Interpret the results of the visual analysis and regression for the third key question: Is the nanosatellite data more or less useful during different seasons of the year (e.g., is the nanosatellite data more or less predictive of same-store sales growth during the busy holiday season for retailers, including November, December and January, or during the non-holiday season)?
  - Interpret the regression coefficient on Holiday*DeltaParkingRate variable.
  - Interpret the P value on the Holiday*DeltaParkingRate variable.
  - Did the results of your analysis align with your expectations?
- Prepare a memo of up to two pages for Maria describing your evaluation of whether the nanosatellite data set can help improve your investment research. Your memo should:
  - Address the potential benefits and risks of relying on the data set.
  - Describe whether the data set is likely to be more useful for research on certain companies vs. others.
  - Consider alternative data sets or technologies that Apollo Investment Research might also consider using.
  - Include appropriate visual evidence to bolster your arguments.

SCORE no. 16691-221US_4