NIKEID DATA ANALYTICS STRATEGY

JOHN BELLAMY

AGENDA

- Summary and introduction to case study
- Data (endogenous and exogenous) & Tooling
- Four pillars of analysis
- Plan of Attack
- Outcomes
- Recommendations
- Conclusion
- About the author

SUMMARY & INTRO TO CASE STUDY

NIKEiD has missed sales targets for the past two months. This team has been tasked with identifying analytical approaches to identify root causes of the drop in sales as well as recommendations to improve decision-making and insight going forward.

To meet the tasks on the left, sample retails sales data was sourced (see next slide) to create concrete examples of recommendations. The repo to support what is demonstrated here can be found in my Github: https://github.com/johnlbellamy/Product-Data-Science

DATA (ENDOGENOUS AND EXOGENOUS) & TOOLING

Because most retail is driven largely by the state of the economy, we can look at both our own internal (endogenous) data and external (exogenous) data. Exogenous data can be data about gas prices, unemployment, the consumer price index (CPI), suppliers or anything else that could affect sales.

Our example data is taken from Kaggle and represents weekly sales data from a retail chain. Information about store type, markdowns and other details are provided. In addition, external data in the form of weekly average unemployment rate, CPI and gas prices are also included to test the hypotheses that these factors have affected sales. The datasets are in my Github and were sourced from Kaggle.

Case study data was viewed and combined and cleaned in Jupyter using Pandas and saved as a csv. This data was further analyzed with data science tools (Scikit-Learn, NumPy, etc.) and loaded into Tableau to create dashboards. Several fields including new calculations Weekly_Sales/Size (sales per sf) and Total Markdowns.

FOUR PILLARS OF ANALYSIS

The four main types of analysis are:

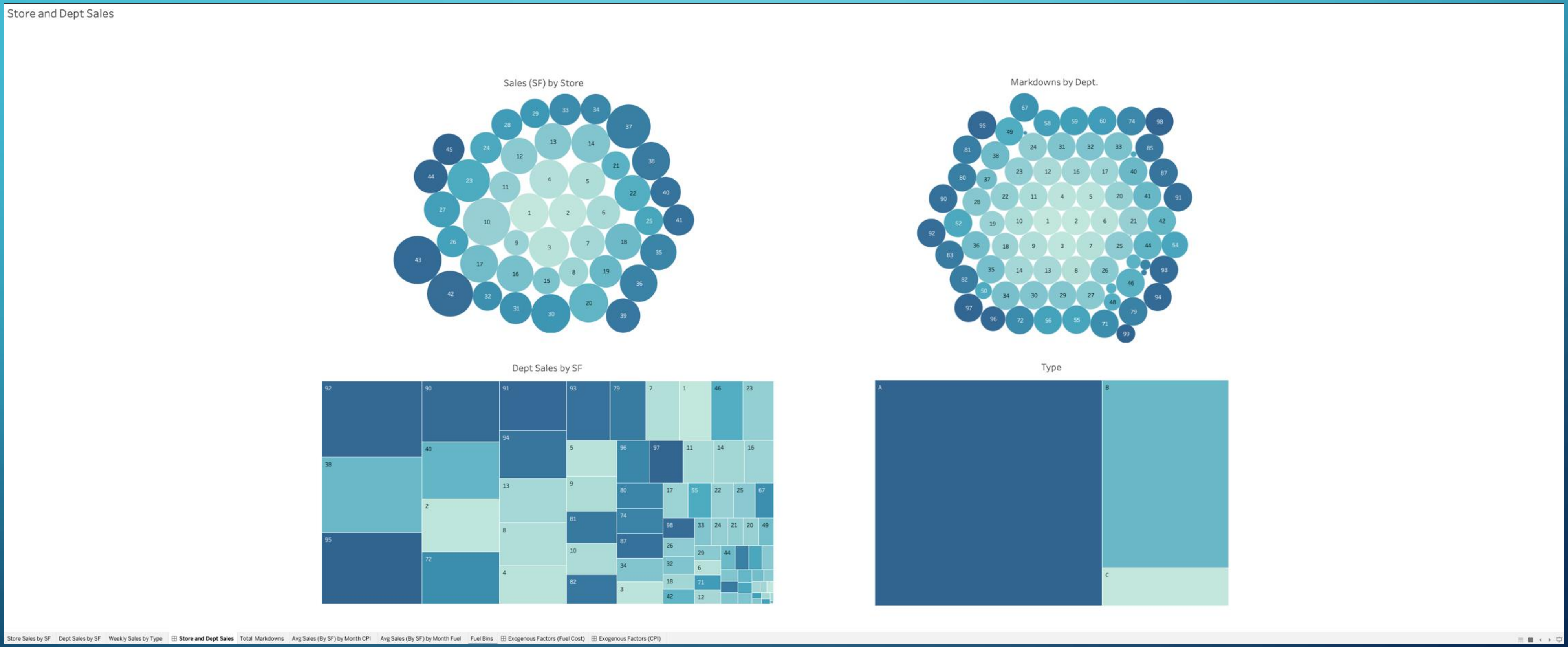
- **Descriptive:** Summarizes and describes past events. Examples are feature analysis and dashboards.
- **Diagnostic:** Examines past performance to find causes. A/B testing. Testing hypotheses about differences in population; high-gas prices result in lower sales, etc.
- **Predictive:** Forecasts future events using historical data and models/ML. What we think of when we say AI. Computer vision, generative AI, recommenders, etc.
- **Prescriptive:** Recommends specific actions based on data analysis. Gauging metrics and adjusting inputs/outputs in response. We see some reality and how do we continue this trend or reverse it.

PLAN OF ATTACK

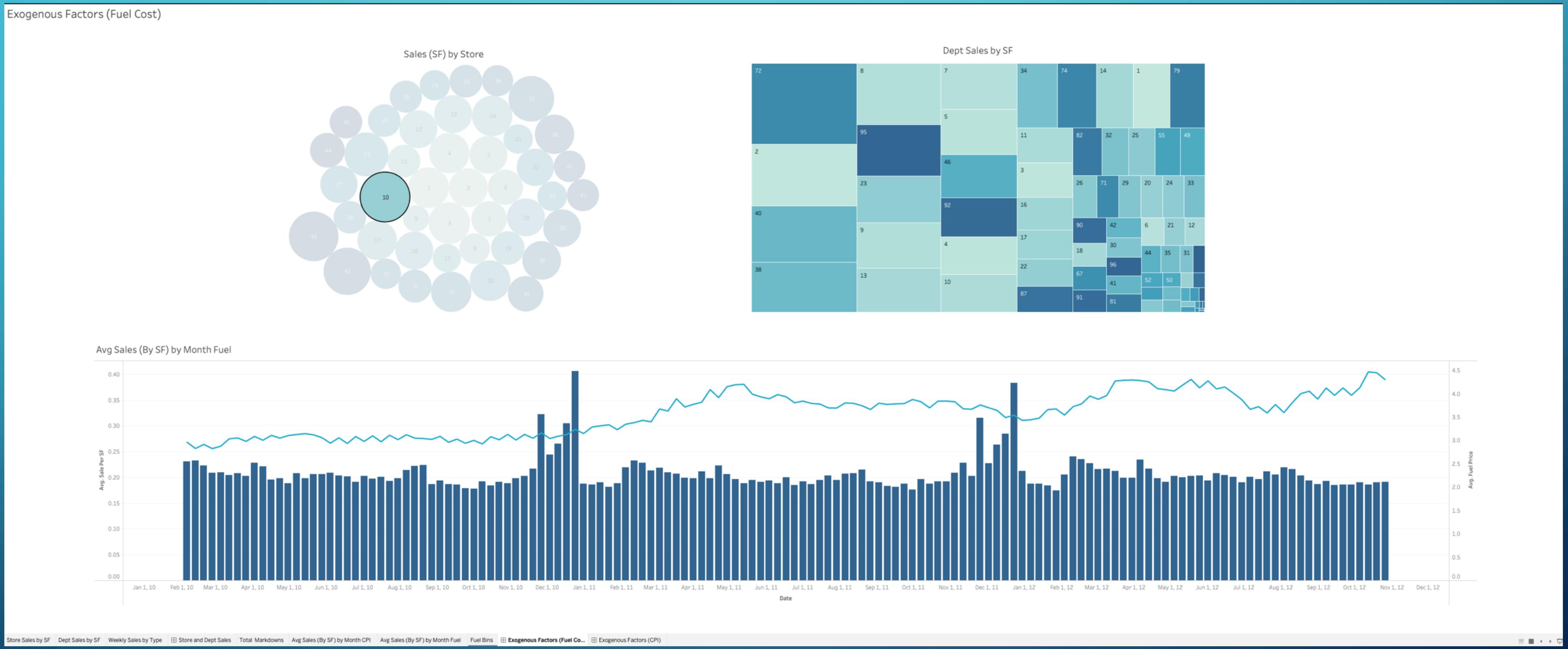
Look at internal and external data points using the 4 pillars to create “Outcomes” as the foundation of recommendations for Nike:

- Descriptive: We can create dashboards to slice and dice data. Are sales down across all venues or only some?
- Diagnostic: Have any changes occurred on the website within last few months and if so test whether these changes might have affected sales. We can also create new A/B testing experiments. We can look at exogenous data and test if external economic/supply factors affected sales.
- Predictive: Recommend NIKEiD shoes to the target demographic and those most likely to buy the brand. Also create a recommender to recommend other products a NIKEiD purchaser might like to increase overall sales. Create new or improve old forecast models.
- Prescriptive: Are some retailers under-performing? Are some regions underperforming? Are some retailers/regions over-performing? We can learn from both cases and try to bring positive outcomes through discoveries.
- The following slides show the outcomes on real data for the four pillars followed by final recommendations.

OUTCOMES (DESCRIPTIVE)



OUTCOMES (DESCRIPTIVE)



OUTCOME (DIAGNOSTIC)

Hypotheses

h0 : There is no difference between the population of u_low_unemployment and u_high_unemployment.

h1: There is a differences between the population of u_low_unemployment and u_high_unemployment.

p = .05

```
# split data into two sets. One where fuel prices are higher and one where fuel prices are lower. A good cutoff could be ~8
low_df = df[df["Unemployment"] <= 8.0]
high_df = df[df["Unemployment"] > 8.0]
```

Executed at 2024.10.09 11:23:44 in 42ms

```
# get random samples representing about 30% of data
x_train_low = low_df["Weekly_Sales_Scaled"].sample(int(round(len(low_df) * .25))) # .25 so we can get somewhat even number of samples from each group
x_train_high = high_df["Weekly_Sales_Scaled"].sample(int(round(len(high_df) * .30)))
```

Executed at 2024.10.09 11:23:44 in 14ms

```
res = stats.ttest_ind(x_train_low, x_train_high, equal_var=False)
```

Executed at 2024.10.09 11:23:44 in 5ms

```
res
```

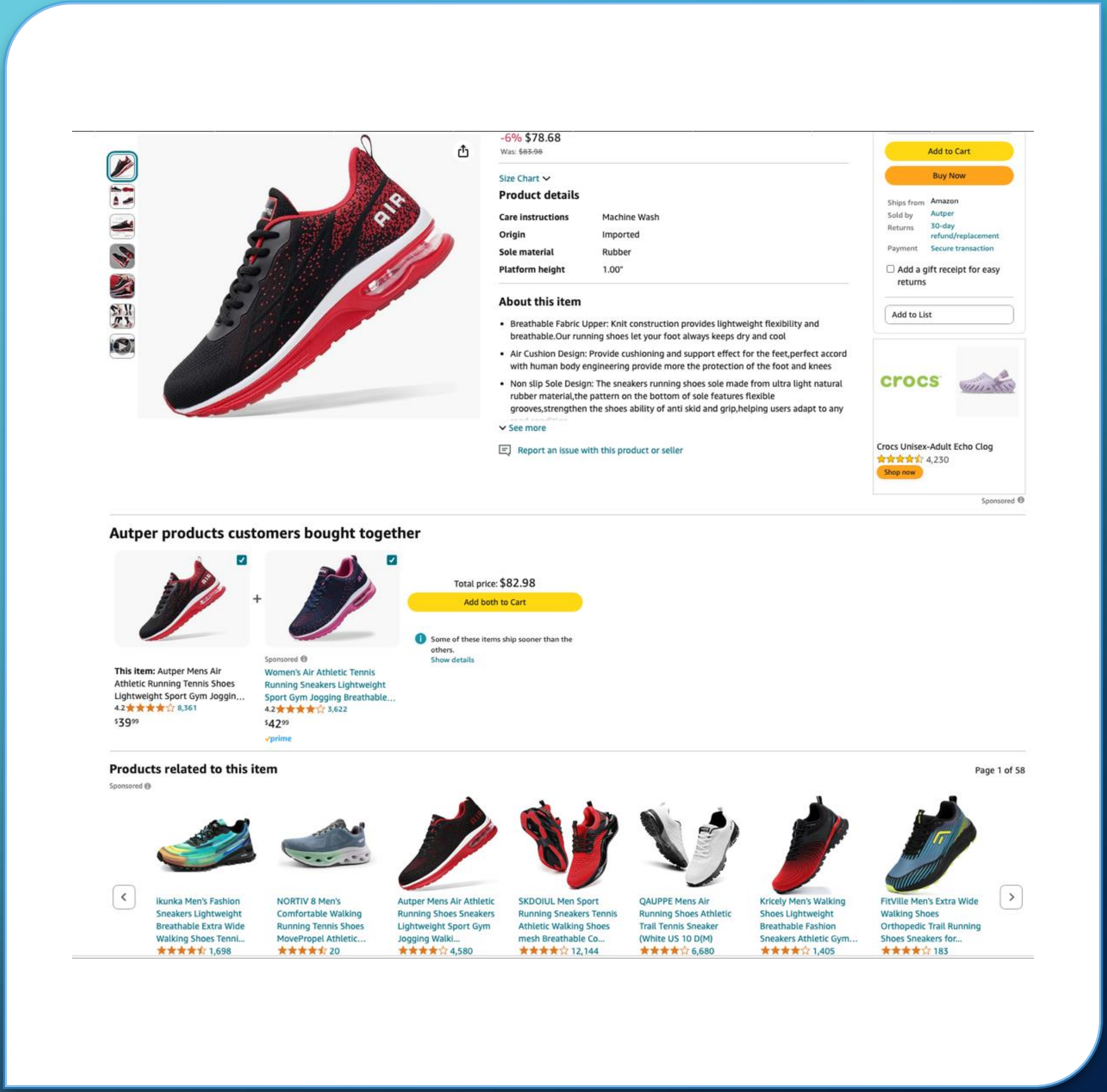
Executed at 2024.10.09 11:23:44 in 3ms

```
TtestResult(statistic=np.float64(-3.203105515794621), pvalue=np.float64(0.0013599845370659708), df=np.float64(96109.11418102641))
```

.0014 < .05 so we must reject the null and accept the alternative.

OUTCOME (PREDICTIVE)

Create recommenders to show potential customers NIKEiD. Also create a recommender to suggest items when users are browsing NIKEiD



OUTCOMES (PRESCRIPTIVE)

- Assume department 92 is the sales leader for this company. A high percentage of sales occurs in this department, yet there are a few stores that aren't selling. Why? What can be done to increase sales?

- We showed that high unemployment affects this company. What can we do to lower costs and pass value down to our shoppers?

FINAL RECOMMENDATIONS

- Check forecast models to make sure they are statistically sound. What are the variables? Have the variables changed? Have assumptions changed? Have external conditions in the world changed? Adjust models, as necessary.
- Develop an A/B testing program to test any website changes creating formal experiments and identifying appropriate metrics. Identify external/internal factors that might affect sales. Test different populations to see if outcomes are statistically due to these factors.
- Make sure current BI program and data engineering program supports dashboarding that will provide the needed descriptive/prescriptive analytics. Create new infrastructure and software, as necessary. Collect appropriate external market and other data to support analysis.
- Evaluate and create new predictive models especially in the areas of product recommendation.
- Finally, when statistical/analytical outcomes occur, make sure all leaders are apprised and able to develop and execute plans to take advantage of opportunities.

CONCLUSION

In this presentation I outlined the four pillars of analysis and provided a case study to provide real and fresh examples from each pillar. I finished the presentation by writing up final recommendations for NIKEiD. My goal was to hit on the major areas of analytics and provide a holistic solution.

Thank you for the opportunity. Had fun.

ABOUT THE AUTHOR

John Bellamy is a highly accomplished Data Scientist and Machine Learning Architect with a strong focus on natural language processing (NLP), computer vision (CV), and ML engineering. Throughout his career he has consistently delivered impactful projects, such as developing GPT-powered information retrieval applications with an impressive 98% accuracy and improving security detection by 40% utilizing Vision Transformer (ViT) models and Bayesian Neural Networks. His expertise spans MLOps practices, where he has reduced model redeployment times drastically, and has led in the areas of automation of machine learning pipelines and XAI.

In his various roles, John has demonstrated strong project management and leadership abilities, including managing multidisciplinary teams, driving new product offerings, and expanding client expenditures. His technical acumen is showcased by his proficiency in key technologies like Python, BERT, Kubernetes, AWS, and more, along with hands-on experience in creating and deploying NLP applications, anomaly detection models, and cloud-native solutions. John's career is marked by a series of key accomplishments, including the successful deployment of Kubernetes clusters in classified environments, reducing product support tickets, and developing scalable ML applications that significantly improved operational efficiency and data-driven decision-making across organizations.