This is a graded discussion: 6 points possible

due Apr 3 at 4:29pm

Data Science Fundamentals and Systems Approaches to Analytics [Videos 1.1–1.3 and Discussion 1.1]

11 94

(https://classroom.emeritus.org/courses/9054/modules/items/1506891)
(https://classroom.emeritus.org/courses/9054/modules/items/1506893)
(https://classroom.emeritus.org/courses/9054/modules/items/1506894)
(https://classroom.emeritus.org/courses/9054/modules/items/1506895)
(https://classroom.emeritus.org/courses/9054/modules/items/1506896)
(https://classroom.emeritus.org/courses/9054/modules/items/1506897)
(https://classroom.emeritus.org/courses/9054/modules/items/1506899)
(https://classroom.emeritus.org/courses/9054/modules/items/1506900)
(https://classroom.emeritus.org/courses/9054/modules/items/1506901)
(https://classroom.emeritus.org/courses/9054/modules/items/1506903)
(https://classroom.emeritus.org/courses/9054/modules/items/1506904)
(https://classroom.emeritus.org/courses/9054/modules/items/1506905)

****There are three videos and one discussion on this page. Please scroll down to view and complete each one.***

Video 1.1: Module Introduction: Data Science Fundamentals (5:11)

In this video, Retsef introduces the teaching approach of this course and the skills you will gain from the upcoming modules. Through the use of in-depth case studies and real-world application

f highlights the connection between technical aspects of data and analytics and the business and organizational considerations.

Video 1.2: Systems Approaches to Analytics and ML (9:12)

It’s not uncommon to view analytics, ML, and artificial intelligence (AI) through a very technical lens. However, this lens is not sufficient in order to leverage these technologies within business environments. In order to do this, a more comprehensive framework is needed. In this video, Retsef introduces a framework that will allow you to think more systematically.

Video 1.3: Data, Models, and Processes (3:39)

The framework that Retsef presented in Video 1.2 relies on three main concepts: data, models and processes. This video takes you through a deep dive into these concepts.

Discussion 1.1: Converting Raw Data Into Usable Data [20 Minutes]

Learning Outcome Addressed:

- Articulate a scenario in which usable data could be overlooked and how best to utilize it.

This is a required discussion and will count toward course completion.

Most of the time, data is difficult to read and understand unless you have prior insights into the data's characteristics. For example, imagine you have a small business that sells baseball caps. You've positioned yourself in the middle of the city, right next to a baseball stadium with lots of foot traffic. Your goal is to create a data model that will predict inventory levels and place orders for more hats before inventory gets low. You configure this predictive model to use previous sales history along with the baseball team's future schedule to predict what sales will look like for the next two weeks. After implementing your model, you find that it's somewhat accurate, but some days are better than others. You hire a specialized data scientist who worked on your friend's food stand's order management system. Through more research, you learn how the weather is a major contributor to your sales. It makes sense that when it's overcast, not as many people want to buy hats to block the sun, doesn't it? However, this piece of data can easily be overlooked when configuring your model. After incorporating the weather outlook into your model, it becomes more accurate, as this key piece of data is now handled as a parameter.

Similar to what Retsef shared in the videos, this scenario illustrates how some data can easily be overlooked and therefore not used effectively in the model. In this discussion, come up with a similar scenario to the baseball hat example or an example from your own experience where data was overlooked and raw data needed to be converted into usable data. Write out the situation on the discussion board, and reply to at least one peer's post with any other suggestion you have to make their model more accurate.

Be sure to read the statements posted by your peers. Engage with them by responding with thoughtful comments and questions to deepen the discussion.

Suggested Time: 20 minutes

Rubric: Discussion 1.1

Criteria	Exceeds expectations	Meets expectations	Below expectations
Thoughtful and complete response to the question(s)	**4 pts** Fully responds to the question(s), post is supported by connections to the reading and real-life examples, and post makes additional connections to the field of data engineering with novel ideas, critical thinking, or extensive application of how to use the topic in future work.	**3 pts** Fully responds to the question(s), and post is supported by connections to the content or real-life examples.	**0 pts** Partially responds to the question(s), or connections to the content are missing or vague.
Engagement with the learning	**2 pt**	**1.5 pts**	**0 pts**

community	Posts thoughtful questions or novel ideas to multiple peers that generate new ideas and group discussion.	Asks questions or posts thoughtful responses to generate a single peer's response.	No responses to peers or posts minimal or vague responses to peers that do not motivate a response (e.g., "I agree.").

Search entries or author

Unread

Subscribed

Reply

(https:/

Yossr Hammad (https://classroom.emeritus.org/courses/9054/users/229118)

Mar 27, 2024

I used to work for a travel organization, and in 2010 we created a trip to the middle east; Egypt. Based on previous sales and requests we decided to purchase many tickets from different airlines. We were sure that all tickets will be sold out in no time. We assumed we have the needed data for our prediction; fascinating country, relatively cheap, beautiful sceneries and beaches. Surprisingly we didnot sell much. when we did more research we figured that we overlooked the political condition of the country; it was unstable for while then a revolution happened.

When we considered the country's condition we changed the destination to different country in Europe and we ended up selling all the tickets.

Edited by Yossr Hammad (https://classroom.emeritus.org/courses/9054/users/229118) on Mar 27 at 6:19pm

Reply

(http

Javier Di (https://classroom.emeritus.org/courses/9054/users/226884)

Mar 28, 2024

That's an interesting case but not sure how you could include a variable for political situation? Maybe a rating 1-10 (10 being worse and 1 great/the US). Adding that variable should improve the predictive power of the model before making the decision to open a route?

Reply (1 like)

(http

Yossr Hammad (https://classroom.emeritus.org/courses/9054/users/229118)

Mar 28, 2024

I am assuming the the political condition falls in the same category as weather changes.
definitely adding that and all variables would improve the predictive power. you have an interesting suggestion of how including that variable.

Reply (1 like)

(http

Lawrence Lumague (https://classroom.emeritus.org/courses/9054/users/225055)

Mar 29, 2024

Hi Yossr,

In your scenario, the country's condition had been affected by social unrest which caused ticket sales in other middle eastern countries to decrease.

Do you think that the raw data that could be examined more is the social opinion regarding the unrest due to revolution and the possibility of a similar situation spreading to neighboring countries in the middle east? It is a possibility that could be factored which people traveling to these other middle eastern countries data on the population's opinion may have been discouraged feeling that they may encounter a similar situation within the time frame of this event.

So, therefore, the elements of raw data that may further be examined could be the social opinion on the population to measure and how they feel the unrest may affect travel to middle eastern countries.

My question is regarding the purchase of the tickets is regarding timing. Did your travel organization purchase the tickets before or during changing political situation? The value of the raw data regarding social opinion on current events could have been a factor towards the decision process to purchase airline tickets within the set of countries in a geographical area.

Thanks for your answer submission.

Reply (1 like)

(http

Dawn Prewett (https://classroom.emeritus.org/courses/9054/users/233112)

Mar 30, 2024

This is a very interesting scenario, thank you for sharing it. I didn't realize that travel agencies would purchase large swaths of tickets and then sell them as part of a package later - though it does make sense now that I think on it. Much like Lawrence, I wondered if the social unrest had started prior to the purchase of the tickets. Is that perchance why the ticket prices were so low or was there other factors that masked that there might be other factors at play? I'm also curious as to what types of changes were made to future models to ensure that these types of factors weren't missed in the future?

Reply (1 like)

(http

Yossr Hammad (https://classroom.emeritus.org/courses/9054/users/229118)

Mar 31, 2024

Hello Dawn,

Thanks for sharing your thoughts.

Surprisingly the tickets price wasnot low, it was as usual price.. the benefit of buying many tickets in advance is that we get whats called (group discount and perks) such as choosing the seats for our clients for free, free ticket or 2 based on how big the group is ...etc.

but as mentioned, the political situation was unstable around countries nearby but we have not considered this variable.

We have decided moving further to expand and have connections to most of the destinations we go to. We opened offices in many countries and allied with travel organizations in other countries. With that, we boosted our sales and avoided any unlikely situation.

Reply (1 like)

(http

Yossr Hammad (https://classroom.emeritus.org/courses/9054/users/229118)

Mar 30, 2024

Hello Lawrence,

Thank you for your response and constructive discussion.

very true what you mentioned regarding spreading the unsettled political situation to other countries which actually was the case.

The answer to your question; my company purchased the tickets before the the revolution or any political changes, however, the neighbors countries had that situation going on.. and we should've considered that situation and avoided any trip to the middle east at that time but we overlooked this important detail which led to a later decision change.

Thank you

Edited by Yossr Hammad (https://classroom.emeritus.org/courses/9054/users/229118) on Mar 30 at 5:33pm

Reply (1 like)

(http

Chris Cosmas (*He/Him*) (https://classroom.emeritus.org/courses/9054/users/226607)

Mar 31, 2024

Hello Yossr,

It saddens me to see that people are afraid to go to Egypt, it has so much to offer (7aga tewga3 2el 2alb).

Have you thought about dividing your customers into customer segments based on their profiles taking into consideration age, travel style, and family situation? It could be that different segments would have different risk appetites, backpackers and younger tourists might be more prone to travel to an unstable area compared to families taking vacations together.

Reply

(http

Yossr Hammad (https://classroom.emeritus.org/courses/9054/users/229118)

Apr 1, 2024

Hello Chris,

People are not afraid anymore.. that was only in 2010 but later on things were stable and we had many trips to Egypt.

We apply this on our marketing plans and we do take in consideration all you mentioned, however, as a travel organization we would be cautious in such situations or similar. even if the client showed interest in going still we cannot be take the risk doing that.

once we learned the situation we had to cancel the whole trip for the tourists, however, we would issue tickets to Egypt to any Egyptian origin clients.

Reply

(http

Ricardo Anaya (https://classroom.emeritus.org/courses/9054/users/228915)

Apr 2, 2024

Know your customer, is as important as knowing your product, great example

Reply

(https:/

Manjari Vellanki (https://classroom.emeritus.org/courses/9054/users/231480)

Mar 27, 2024

I would like to illustrate a scenario where data can be overlooked and raw data needed to be converted to usable data.

As I'm working in a Clinical domain, as part of request we are analyzing how the PK parameters of individual patient will be affected by drug concentrations. For that we have raw data variables like subject id, Blood sample collection date/time for PK (like actual timepoint and nominal time point), concentrations for different parameters and so on. But while trying to generate the Tables / Figures, not getting the appropriate results. While the team is trying to figure out the issue and find out that we have overlooked one variable that is difference in time between the Actual collection time point and Nominal collection time point. If the difference is relatively high we should not include those records into consideration. And we got appropriate results after excluding those records as part of analysis.

Edited by **Manjari Vellanki (https://classroom.emeritus.org/courses/9054/users/231480)** on Mar 27 at 5:34pm

Reply

(http

Javier Di (https://classroom.emeritus.org/courses/9054/users/226884)

Mar 28, 2024

Interesting case. This one sounds more like cleaning up/excluding data rather than having missed a key variable in the model to make it more predictive? I have no expertise in the field so may not have understood it in full. Thanks

Reply

(http

Manjari Vellanki (https://classroom.emeritus.org/courses/9054/users/231480)

Mar 28, 2024

Hi Javier,

Thanks for your response. The scenario actually involve in overseeing a key variable that collects the information about the difference in time between actual collection and planned collection of samples which results in showing inaccurate results.

Reply

(http

Roy Nunez (https://classroom.emeritus.org/courses/9054/users/229552)

Mar 30, 2024

In this scenario, would it be more helpful to exclude the nominal time points rather than exclude the records that have high differences?

Or if nominal time is a critical data point, could the difference between the nominal and actual become its own parameter with a weighted score?

I am not sure what the degree of importance does the nominal time have, but I am thinking of another way we could not exclude records that could be vital.

Reply

(http

Manjari Vellanki (https://classroom.emeritus.org/courses/9054/users/231480)

Apr 1, 2024

Hi Roy-

Yes, nominal point(Planned time point as per Study protocol for PK Blood sample collection). As Clinical studies are critical, and are well designed to achieve robustness and chances of overseeing any data is minimal. While we receive the data as raw data, our department is involving in activities like data cleaning, preparing datasets ready for analysis. For example, if the programming team has to generate a table or graph out of Analysis dataset, we'll have apply few criterion flags to subset the original data and perform the statistical methods. while applying criterion flags there is chance of overseeing few key variable flags(reasons might be mis understanding of requirements, poor programming knowledge..).But we perform multiple runs on same program and after being thoroughly reviewed by team, we'll finalize the result.

Reply

(http

Roy Nunez (https://classroom.emeritus.org/courses/9054/users/229552)

Apr 2, 2024

Manjari,

Okay, thank you for helping me understand better. Still wonder weighting them based on time difference, like the longer the difference a lower percentage. I was reminded again of weighing parameters in the Optima and Quanta case Retsef discussed. Maybe the team has already thought about this and it does not apply or help at all.

Reply

(https:/

Chris Cosmas (*He/Him*) (https://classroom.emeritus.org/courses/9054/users/226607)

Mar 27, 2024

In my previous experience at a financial data analysis firm, the work environment was highly process-oriented. As it was a large well-established organization, processes were standardized and well-defined, with each process consisting of various tasks that had to be completed within specific timelines (quarterly, weekly, daily, hourly, etc...).

One particular process involved reviewing legal documents provided by financial firms and extracting specific data points. During two quarters, our team encountered challenges in meeting the timeliness metric for completing these tasks. This was highlighted by key performance indicators (KPIs) used to track our progress. During a brainstorming session with my peers, we came up with the idea to develop a function based on Linear Programming in Operational Research. The function would take different weights such as languages spoken, number of colleagues, amounts of tasks at the beginning of the day, the average time it would take for a colleague to finish tasks and other variables. The primary objective of this function would be to distribute tasks among team members in the most efficient manner possible, minimizing the impact of all variables involved to ensure optimal task completion. We ran the program several times to test the results but still got complaints from team members. We realized we had overlooked a very important variable which is the different titles of each of my colleagues, as team members ranged from junior analysts to senior analysts, each role posed different responsibilities with more senior roles having larger and more time-consuming tasks, additionally, my colleagues were involved in different projects each varying in complexity. It was difficult to quantify the time spent on other tasks for each team member but a colleague of mine assigned a new variable with different weights for all team members based on seniority and process ownership (the number of other processes handled by each team member), this improved the task allocation of the function. Using the algorithm not only allowed us to eliminate a process that required a senior team member to allocate tasks manually daily but also ensured equitable and fair task distribution.

Reply

(http

Ahmad Abu Baker (https://classroom.emeritus.org/courses/9054/users/234460)

Mar 27, 2024

Your experience at the financial data analysis firm highlights the importance of leveraging data-driven solutions to optimize task allocation and improve operational efficiency. Your journey from encountering challenges in meeting timeliness metrics to developing and refining a Linear Programming-based function demonstrates critical thinking and problem-solving skills.

Your approach of using a function based on Linear Programming to distribute tasks among team members efficiently is commendable. It not only addressed the immediate issue of task completion but also minimized the impact of various variables involved, such as languages spoken, number of tasks, and average task completion time.

The pivotal moment in your story comes with the realization that the initial function overlooked an essential variable – the different titles and responsibilities of team members. This insight led to the creation of a more refined algorithm that considers seniority, process ownership, and task complexity when allocating tasks. This adjustment not only improved task allocation but also ensured fairness and equity among team members.

Your experience underscores the significance of continuous improvement and adaptation in data engineering and operational research. By incorporating feedback, refining algorithms, and considering diverse variables, you were able to streamline processes, eliminate manual tasks, and achieve optimal task distribution.

Furthermore, your story aligns with the broader principles of data engineering, emphasizing the use of algorithms, optimization techniques, and data-driven decision-making to enhance organizational performance. Your innovative approach and problem-solving mindset serve as an excellent example of how data engineering concepts can be applied effectively in real-world scenarios to drive efficiency and productivity.

In future work, building on this experience, you can explore advanced optimization algorithms, machine learning models for task prediction, and automated task allocation systems to further enhance operational efficiency and resource utilization within the organization. Your experience serves as a valuable lesson in leveraging data engineering principles for continuous improvement and achieving organizational objectives.

Reply

(http

Turki Alghusoon (https://classroom.emeritus.org/courses/9054/users/229165)

Mar 27, 2024

Hi Chris,

Interesting example and a smart move to incorporate seniority and ownership into the model.

I wanted to share a thought I had while reading your story:

It might be useful to track the team members' performance on tasks that were assigned to them in the past and use that as an additional variable into the model. This could be implemented by analyzing the historic performance at the different levels of project complexity, aggregated at the seniority level. Proxies for performance could be average completion time, or delay as % of original timeline.

Reply

(http

Javier Di (https://classroom.emeritus.org/courses/9054/users/226884)

Mar 28, 2024

Agree with Chris and this seems like a great case for machine learning and feeding the data to optimize performance of tasks allocation amongst team members and removing the discretionary/subjective role of a staffer without using data

Reply

(http

Chris Cosmas (*He/Him*) (https://classroom.emeritus.org/courses/9054/users/226607)

Mar 29, 2024

Hello Turki,

Thank you for your reply.

We did take into account the average time used by my team members but another issue had risen. Many tam members presented pushback on the useage of their performance data on the model. They were hesitant to reveal that they were slower in comparison to the team or made more mistakes compared to the team average. This brings up another discussion altogether which is the ethics of using some of the data available to decsion makers. From a managerial perspective it was useful for our Team Leader to take the data into account as he already tracked it for personal use, but when the discussion came up on our weekly meeting many team members were unhappy. Team mebers communicated this was personal information which should not be disclosed with all the team. This issue could also be due to cultural differences as cultures tend to value privacy and personal rights differently.

Reply

(http

Turki Alghusoon (https://classroom.emeritus.org/courses/9054/users/229165)

Mar 30, 2024

That makes sense. We ran into a similar issue with staff pushback on use of individual performance data for analysis. We managed to get buy-in from the teams by explaining how the data was only going to be used in aggregates with the sole objective of achieving overall process improvement. We kept hammering in that the data was not going to be used to evaluate team members. Eventually we were able to get the buy-in but I fully understand that every situation is unique.

You bring up a good point about the ethical use of data. At the end of the day, data in situations like these does represent actuals human beings, and due care should be given.

Reply

(http

Haitham Farag (https://classroom.emeritus.org/courses/9054/users/233864)

Apr 2, 2024

Good day Turki

The use of Individual performance data (preferably objective data) for analysis, is such a tricky subject to introduce in many organizations. I have seen it more successful when it's grounded in the organisational culture, and its intended use, to develop rather than punish is effectively communicated.

Still, this is easier said than done. KPI's and balanced scorecard concepts have been around for decades but the level of adoption is not proportionate given the value they can bring.

It is always reassuring to read about other peers' similar experiences.

Thanks for sharing

Edited by Haitham Farag (https://classroom.emeritus.org/courses/9054/users/233864) on Apr 3 at 6:58am

Reply (1 like)

(http

Manjari Vellanki (https://classroom.emeritus.org/courses/9054/users/231480)

Mar 28, 2024

Hi Chris,

I really liked the idea, implementation and how your team overcome the issue that has been overseed. We usually spent hours in monthly meetings for my Manager to assign tasks manually and we are ending up with work load imbalance across team members. I like the thought of assigning tasks based on team dynamics which is one of the best strategy of good leadership.

Reply

(http

Swati Sharma (https://classroom.emeritus.org/courses/9054/users/236938)

Apr 3, 2024

Hello Chris: Thank you for sharing! I'm curious about the timeframe it took to develop the algorithm. Additionally, once it was completed, how did you assess the situation if a team member couldn't meet their assigned task deadline? Did they not assigned new tasks until completion?

Furthermore, i am curious if you had implemented a task performance rating based of the task completion, would have enhanced team motivation and helped the manager to understand the roadblocks in front of the team members.

Additionally, how are the following things evaluated in the model?

1. How was vacation/time off/sick leave considered by the model when assigning the task?

2. How was the skillset aligned to the task assignment to an employee? I see that you had considered the title of the employee in the model but would skillset be mapped to the title of the employee? Something to consider as part of the model development.

Reply

(https:/

Ahmad Abu Baker (https://classroom.emeritus.org/courses/9054/users/234460)

Mar 27, 2024

In my role within a government authority's Project Management Office (PMO), I encountered a situation that emphasized the criticality of comprehensive data utilization in baseline setting. We were tasked with overseeing a major infrastructure development project aimed at enhancing public transportation in a metropolitan area.

Initially, our focus during the baseline setting process was primarily on budget allocations, project timelines, and resource planning. However, we soon realized that overlooking certain data elements could have significant implications for the project's success.

One key aspect we initially underestimated was the impact of regulatory approvals and compliance requirements on the project timeline. While we had accounted for standard permitting processes, we hadn't fully considered potential delays due to unforeseen regulatory changes or stakeholder consultations.

To address this challenge and improve our baseline accuracy, we implemented the following strategies:

Firstly, we conducted an in-depth analysis of historical data from similar infrastructure projects, specifically focusing on regulatory approval timelines, public consultation feedback, and legal compliance challenges. This data-driven approach helped us anticipate potential bottlenecks and adjust our baseline accordingly.

Secondly, we enhanced stakeholder engagement by collaborating closely with regulatory agencies, local authorities, and community representatives early in the baseline setting phase. Their input and insights allowed us to identify regulatory risks, streamline approval processes, and align project milestones more realistically.

Thirdly, we integrated risk management practices into the baseline setting process, identifying and mitigating potential risks related to regulatory changes, permitting delays, and stakeholder disputes upfront. This proactive approach helped us address regulatory hurdles proactively and minimize their impact on project progress.

Lastly, we established a robust monitoring and evaluation framework to track regulatory compliance, milestone achievements, and budgetary expenditures throughout the project lifecycle. Regular reviews, progress assessments, and data-driven decision-making enabled us to adapt to changing regulatory environments, optimize resource allocation, and ensure alignment with project objectives.

By leveraging historical data, enhancing stakeholder collaboration, integrating risk management practices, and implementing a robust monitoring framework, we significantly improved baseline accuracy and successfully navigated regulatory challenges within the government authority's PMO. This experience underscored the importance of data-driven decision-making and proactive risk management in ensuring the successful execution of large-scale infrastructure projects.

Reply

(http

Dawn Prewett (https://classroom.emeritus.org/courses/9054/users/233112)

Mar 29, 2024

This is definitely a common issue in project management, but seeing the whole picture can be difficult, especially when it isn't always clear what the whole project is. Iterative phasing can help with this, but your stakeholders do not always give you that berth. What triggered the realization that overlooking certain data elements could have significant impact on a project's success and how did you identify those elements that should have been considered all along? You also mention robust monitoring. What exactly was being monitored and what thresholds did you put in place? Did you ever have to take action and if so, what kind of action did you take to get the project back on track?

Reply

(https:/

Haitham Farag (https://classroom.emeritus.org/courses/9054/users/233864)

Mar 27, 2024

Case: **Refugee Camp Design**

Objective- the "Business" decision to be improved

Address the questions regarding *where* and ***how*** many ***water point facilities*** are to be constructed to meet the needs (as per set international standards) of the influx of refugees into a camp setting.

"Business"/Operational Context

The influx of over one million Rohingya refugees in Cox's Bazar (Bangladesh) in 2018

Available Raw data

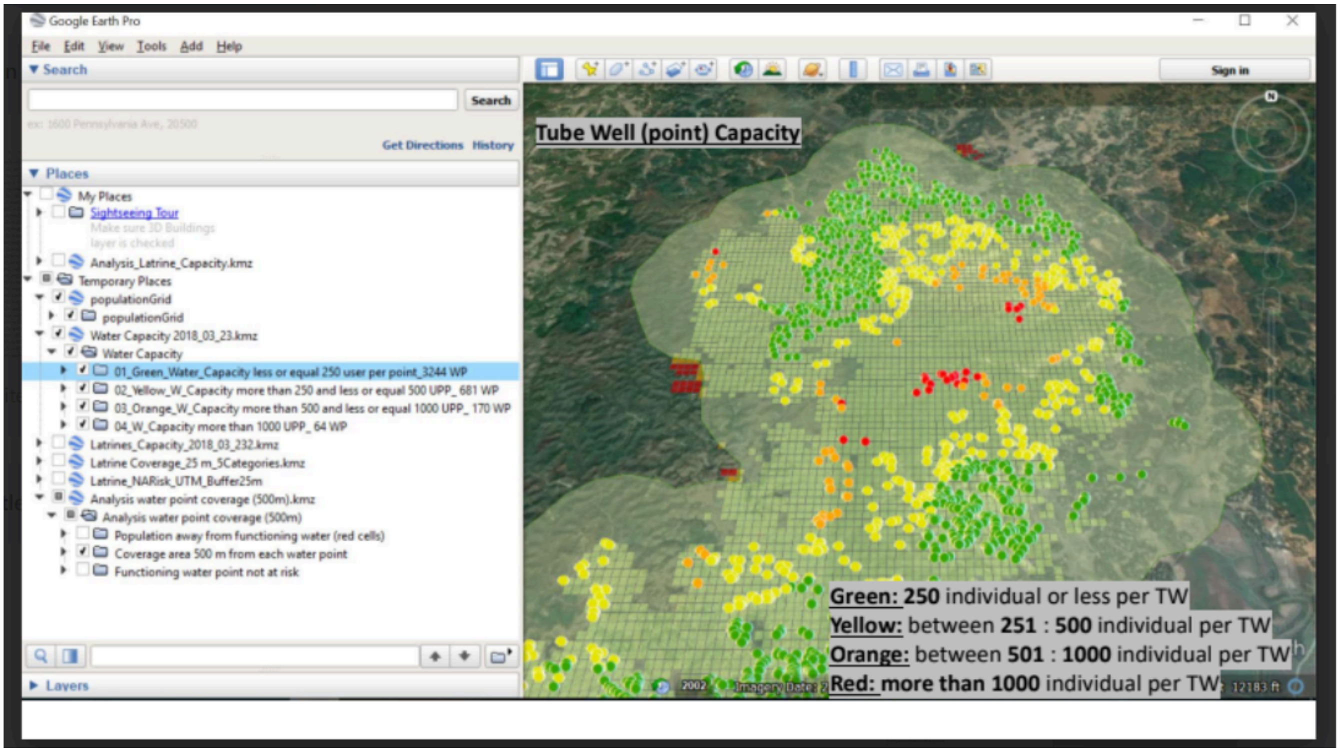
-Geospatial data and census of

- refugees (by gender and age)
- existing operational water points

-natural risk susceptibility maps

Usable Data

Raw data was processed to produce detailed maps of ideal locations to drill for water (establish water points) to cover unserved and underserved refugee households. Coverage is defined as per the international standards (i.e. number of people served per water point and distance from household to the nearest water point)



Overlooked data

During the rainy seasons and the occasional cyclons, cause of flooding and landslides resulted in water points being

- inaccessible access (flooding around)
- washed away or covered by flood debris
- contaminated by the floods

Outcome/ Lesson Learned

Not accounting for flood and landslide terrains, and not transposing both the existing and suggested water point location on the available (but unutilised) *natural risk susceptibility maps*, resulted in overcrowding at the water points that were still operational after the rainy season.

Edited by **Haitham Farag (https://classroom.emeritus.org/courses/9054/users/233864)** on Mar 27 at 11:18pm

Reply

(http

Turki Alghusoon (https://classroom.emeritus.org/courses/9054/users/229165)

Mar 27, 2024

Hi Haitham,

Cool project and a noble cause! I had 2 questions related to the available data:

- Did the geospatial data include elevation information?
- If elevation data was available, could that be combined with the risk map to determine optimal locations for water points (where the points could be at elevations high enough as to not be impacted be floods?

Reply

(http

[Haitham Farag (https://classroom.emeritus.org/courses/9054/users/233864)](https://classroom.emeritus.org/courses/9054/users/233864)

Mar 28, 2024

Thanks for the feedback Turki.

The campsite elevation information was accounted for in the landslide hazard and flooding risk maps, which unfortunately were not initially used. Your point is spot on, elevation impacts both accessibility (walking 500 m to a water point on a flat vs, hilly terrain) and sustainability of access to water. Moreover, some of the hills had the risk of landslides.

A costly example of *hindsight is 20/20!*

Reply (1 like)

(http

[David Taylor (https://classroom.emeritus.org/courses/9054/users/233381)](https://classroom.emeritus.org/courses/9054/users/233381)

Mar 28, 2024

Wow this is definitely a good example! It makes a lot of sense that rainy season could impact the availability of the wells. My immediate thought was how difficult it would be to estimate the potential impact of rain, and so it's quite excellent that there exists such a thing as *natural risk susceptibility maps.* I wonder what goes into those maps, elevation clearly but I wonder if they also account for the type of ground/dirt/rock/porosity of a given location.

Reply

(http

[Haitham Farag (https://classroom.emeritus.org/courses/9054/users/233864)](https://classroom.emeritus.org/courses/9054/users/233864)

Mar 31, 2024

Thanks for the feedback, David.

Geologists tested the grounds, and I believe ground sampling and maybe seismic surveys were conducted pretty much like what you have pointed out. The maps were not perfect cause deforestation of the hills to make space for camp exasperated the landslide risk. The implication of heavy deforestation is probably another missed data!

Interestingly, after the first year in the camp refugees having experienced the same weather conditions in similar terrains started introducing makeshift and sometimes resilient solutions to some of the challenges caused by flooding.

Reply

(http

Lee Lanzafame (https://classroom.emeritus.org/courses/9054/users/231975)

Mar 30, 2024

great example, as someone who was recently flooded, we consider the min/max height previous floods have gotten to but we also consider something called the 1:100 year flood, which is the statistical likelihood of (1 in 100) 1% probability of a flood happening in any given year. In Australia it's a regulatory benchmark for floodplain management and building design. great job

Reply

(http

Haitham Farag (https://classroom.emeritus.org/courses/9054/users/233864)

Mar 31, 2024

Thanks, Lee

Ground saturation also seems to play a huge part in flooding. I echo, your point, regulatory policies are proven to be necessary.

Sorry to read about you experiencing flooding. I hope you have fully recovered from the flood's aftermath.

Edited by Haitham Farag (https://classroom.emeritus.org/courses/9054/users/233864) on Apr 3 at 7:06am

Reply

(https:/

Turki Alghusoon (https://classroom.emeritus.org/courses/9054/users/229165)

Mar 27, 2024

A few years ago, I was tasked with determining the main factors impacting the elapsed time of audits (in calendar days), in order to optimize delivery time for audits in the annal portfolio. Data initially used in the analysis included:

- # of stakeholders involved,
- complexity of audit area (1-10 scale; where 10 is most complex)
- # of systems supporting the audit area
- extent of data analytics support required for the audit (high, medium, low)
- # of country office visits required for the audit

After preparing the data and experimenting with the initial model, I found that the data points used did not have a significant impact on predicting the elapsed time for audits. For example: While the "# of country-office visits required" had positive correlation with elapsed time for audits, it did not have any impact on predicting the elapsed time for audits that did not involve any country-office visits.

I then decided to look for other factors that could have an impact on the elapsed time and decided to examine the time of the year when audits are launched. I observed the following:

- Audits that started between (May-July) and audits that started between (October-December) had a longer average elapsed time than audits starting in other months of the year.
- The elapsed time was further extended if the audits involved a large number of stakeholders.

Upon sharing the observations with the auditors, they explained that delays in those audits were mainly due to the unavailability of important stakeholders who were on leave (summer breaks, Thanksgiving and Christmas holidays). This was a major insight that I was able to share with the leadership team. As result, the unit changed the scheduling approach so that more complex audits that have a large number of stakeholders are scheduled to start in Q1 (January-March) or later in Q3 (August – September) in order to minimize client availability during critical stages of the audits.

Potential refinements: If I had access to anonymized leave information for all staff in the organization, I would be interested in running the following experiment:

1. Import and aggregate leave data for all staff within the organization.
2. Re-calculate leave as a percentage of staff availability for each month across the organization.
3. Test the impact of the new variable (defined in step 2) on the elapsed time for audits.

Reply

(http

MATT DEFREITAS (https://classroom.emeritus.org/courses/9054/users/220100)

Mar 28, 2024

Turki, based on your post it sounds as if you had a rigorous process of analyzing the factors that would impact audit elapsed time. This process eventually led to some actionable insights for the business which aligned perfectly with these videos.

As we must think about this from a business lens as well, is anonymized data something HR would sign off on? Or do you anticipate any challenges obtaining such data? In some companies they blackout leave during their busiest periods of time to avoid any of these pitfalls. Has that ever been discussed as an option?

Reply

(http

Turki Alghusoon (https://classroom.emeritus.org/courses/9054/users/229165)

Mar 28, 2024

Hi Matt,

Thank you for sharing your thoughts. Obtaining leave data for the whole organization is quite tricky as you might have guessed even when anonymized. That is due to the HR policies in place which are deigned to protect the privacy of the staff; and although such information could be released in business-critical scenarios, schedule optimization at the business-unit level is not considered business-critical.

As for leave blackout period, it has never been discussed as an option, although I believe it would significantly improve efficiency during crunch periods.

Thank you for your feedback!

Reply

(https:/

MATT DEFREITAS (https://classroom.emeritus.org/courses/9054/users/220100)

Mar 28, 2024

In a sales environment, our client was leveraging the number of sales (fixed cost) to forecast their total sales. One month into the quarter, their forecast began to decline significantly. They engaged our team to try and uncover the problem as this decline meant they were going to be significantly under their target. Once we were able to understand their approach, we quickly understood the situation. We uncovered that the problem was they weren't really forecasting but rather had developed a pacing tool. By basing their forecast on the number of cumulative sales (an output) against the prior year's sales, they were making too many assumptions about their processing times (inputs).

Once we dove into the data, we quickly discovered some issues with their inputs which would result in a decline in sales:

- The sales cycle time had been lengthened meaning more customers were taking longer to purchase. This alone meant that they were not going to be able to keep up with prior year's sales numbers as the timing was not the same.
- We uncovered some staffing shortages in one of the most critical phases of their sales process. This shortage was creating a backlog of applicants and would have a severe impact on the timing of the sale.

As Retsef alluded to in the opening video, we must think end-to-end about the process in a business perspective. The forecast was shortsighted because it was only using the end result. Someone might have all the technical skills in the world but if they are not able to apply that knowledge to the business domain the results may not be as successful.

We did just as Retsef mentioned in his second video referring to data interpretation. We conducted a deep dive analysis on their processes holding stakeholder interviews and speaking with technical experts about the available data points. We needed to fully understand not only the business process but also how the data was recorded so that we could either leverage as is or make the data more usable.

A model was developed leveraging the sales cycle starting with marketing spend, mix, progression, and sales duration timing resulting in a forecast that was 97% accurate. This outcome was achieved we took the time to understand the life cycle of the sale and brought forth the inputs that would influence and/or effect a sale.

Reply

(https:/

Roman Jazmin (https://classroom.emeritus.org/courses/9054/users/225803)

Mar 28, 2024

Initial Statement:

Based on what initial factors that you know about or consider relevant, these things will shape the outcome of your initial predictive model used in one's operations. Discovering and considering all factors in a given operations depends largely on performing A/B test using sample data (I would have the initial dataset be a mix of historical and present data), running the data to an initial predictive model, study its initial outcome or results, make or discover new factors to consider in our business process, and then update our predictive models based on what we learned.

Run another test to see if our outcome improves. We continue to perform this task until management agrees that we have reached the best possible outcome, having discovered, and applied all possible factors or input at that point and time.

Scenarios:

Let us start talking about possible scenarios. For example, teams not performing as well as in past years. One must consider the stock market statement, “Past performance doesn’t guarantee present or future success or outcomes.”.

Team performance does affect sales like the cost of stockpiling inventory for one given team over another. A team’s schedule for the season is not only a factor to consider, but one must also consider the teams’ present roster.

Not being able to predict a team’s roster for the upcoming year (i.e.. draft picks). A team’s performance depends primarily on each member’s performance. Some hidden factors to consider include:

1. Each person’s performance differs significantly based on field position, experience levels, and amount of previous training.
2. A person’s feel for the new environment.
3. How a player performs or works in a little league environment differs at a high school, college, and professional levels. Again, historical stats do not guarantee present or future performance or outcomes.

A more personal example to me is when I was working as a data scientist for a domestic, U.S. company that had an international client based on Europe. The task was to find out which customer-related issues, mapped out to its corresponding answers or solutions, for a particular product will be asked to a Customer Service Representative responding to calls.

One possible missing factor is:

1. Demographics –
2. Cultural preferences based on habits.
3. Age group concerns
4. Different Age groups have different concerns than other Age groups living in different areas.

The differences can be in terms of size, color, material, brand that a particular age group prefers as opposed to the same age group living in another part of the world.

1. Cultural buying habits:
2. What time of the day do most people in the area buy online?
3. Timeframe that people who do to work, are off work and then order online.

A1. Determine which days or seasons are best-selling times for a given product for a given country or area in the world.

1. People who work from home and time that they go online and order online for a given product for a given country or area in the world.
2. Which ads generate a buy for a given culture or people group.
3. Which words attract that leads to a sale.
4. Which images attract that leads to a sale.
5. Which colors used in the ads attract that leads to a sale?

Conclusion:

Constant testing of a given dataset followed by updating the present predictive model for an operation are keys to success. Change is always happening so companies must be willing and able to pivot to have its predictive accurate and relevant.

Reply

(http

Mhelissa Yayalar (https://classroom.emeritus.org/courses/9054/users/233590)

Mar 29, 2024

Hi,

I shared similar perspective with predictive analysis on retail sales, but you brought up another data input that I didn't consider, which is demographic data of the consumers.

In my example, the Christmas niche business is catering to a particular consumer and overlooking buying habits of consumers who do not celebrate Christmas holiday may introduce anomalies in my models. Another factor that I also think about is the product offerings to offer particular demographics that does celebrate Christmas. For example, purchasing name tags like, "Noel, "Feliz Navidad" may cater to South American cultures.

Again, to your point, accounting for the demographic data of the retail consumers can increase the accuracy of the predictive models.

-my

Reply (1 like)

(http

Roman Jazmin (https://classroom.emeritus.org/courses/9054/users/225803)

Mar 30, 2024

Morning Mhlissa,

Thank you for your reply. I used to have an eBay account/store. A few things I realized that were hidden data sets were what time of the day sales for a particular product is up and if they are workers that drive to work as opposed to those who work from home. It seems like those who have to drive to work everyday are least likely to buy a particular product on a normal week. But on holidays it seems everyone is online and looking for deals. Another thing to consider is how popular the latest version of a product is as compared to its previous version. People don't want to pay an arm or a leg just to get the latest version of a product. That has been my experience.

Cheers,

Roman

Reply

(http

Haitham Farag (https://classroom.emeritus.org/courses/9054/users/233864)

Apr 2, 2024

Good day Mlhelissa

Your example provides much food for thought and reflection on how it could apply to other sectors (Cash transfer to the most in need as a livelihood coping mechanism). In development, the tendency is to collect much more data than what is used. Being able to look at the demographic data and see its potential is an area with much room for improvement.

Thanks for sharing this dimension of looking at data.

Reply

(https:// [Javier Di (https://classroom.emeritus.org/courses/9054/users/226884)](https://classroom.emeritus.org/courses/9054/users/226884)

Mar 28, 2024

In my own investing experience over the past few years and since interest rates were kept at near zero rates after the Great Financial Crisis in 2008, Technology stocks have greatly outperformed other sector investments. Most investors attributed this to the great demand for technology products and their likely revenue growth.
Running a data regression between the Technology sector companies revenue growth and stock performance may be somewhat accurate but not complete as it would be missing a key factor to understand. This key factor missing is that interest rates have been very low through the period and Tech stocks benefit and get revalued upwards in periods of low interest rates, given that most of their earnings and growth are in the distant future.

Incorporating interest rate as a relationship driving Tech stocks performance significantly improves the predictive power of the model, making it more accurate, as this key piece of data is now handled as a parameter besides looking at business revenue growth.

As evidence by the inverse correlation below for Tech Stocks Performance & Interest Rates

Edited by **Javier Di (https://classroom.emeritus.org/courses/9054/users/226884)** on Mar 28 at 6:36pm

Reply (1 like)

(http **David Taylor (https://classroom.emeritus.org/courses/9054/users/233381)**

Mar 28, 2024

Javier this is a great example. Do you think the main reason for this is "that most of their earnings and growth are in the distant future"? A good comparison would be to run the same analysis on another industry whose earnings are more immediate and/or are very slow to grow.

Reply (1 like)

(http **Javier Di (https://classroom.emeritus.org/courses/9054/users/226884)**

Mar 29, 2024

Thank you David, that is correct and the reason for the relationship with interest rates. If you run the same analysis on value stocks that would trade at a low multiple of free cash flow it would show a much lower correlation and why value tends to outperform growth in times of very high interest rates. Hope this helps

Reply (1 like)

(https:/ **David Taylor (https://classroom.emeritus.org/courses/9054/users/233381)**

Mar 28, 2024

Working for a database marketing company, I feel like we run into these scenarios all the time. One such situation was related to measuring the success of email marketing campaigns and predicting who would respond the best (ie buy something). Generally a consumer will get a series of marketing emails, and the client always wants to know what impacts responsiveness the most. The content itself is of course a major factor. One might think it is the primary factor. However, one employee thought to ask, "Let's break it down by several another metric. How does # of emails sent impact the outcome?"

By breaking it down in that way we are able to convince our clients to send emails more often because X more emails will convert Y% more customers. And yet another missing piece of the puzzle was physical mail (like an actual post office). Turns out customers are way more likely to bite if they get physical mail and email. Not only that, but if they are timed so that they arrive around the same time, the effect goes up again.

What could other factors be that influence an email campaign?

Reply (2 likes)

(http

Timothy Andrew Ramkissoon (https://classroom.emeritus.org/courses/9054/users/226697)

Mar 29, 2024

David, your question is quite broad, I'll stick to the effectiveness of digital vs physical emails.

- How are you tracking whether or not the recipient(s) of the email received said email? For digital emails, there is the possibility that it could've ended up in their spam/junk folders, especially if they have a corporate email. Similarly for physical mails, these can be lost/damaged in the process.
- Does increasing the quantity account for this shortcoming? Is the increase in quantity able to increase visibility in some regard? For physical mails, does quantity overcome the challenge of loss/damaged mails.
- Is there a turning point where quantity of mails sent change from opportunity to threat, where the recipient changes their opinion from positive to negative?

These are some factors I would also consider when building my business strategy for this senario.

Reply (2 likes)

(http

David Taylor (https://classroom.emeritus.org/courses/9054/users/233381)

Mar 29, 2024

Yes in fact too many emails does turn to a negative outcome!

As far as the tracking goes, the physical mail is quite limited. However for email, generally if it is marked as spam by the email service or is undeliverable, we are notified and we don't count that as delivered. Beyond delivery, we can also tell if someone opens the email, and finally of course if they click the link in the mail then we have another data point. So we will breakdown the lead rate by delivered, opened, and clicked rates.

Reply (1 like)

(http

[Jignesh Dalal (https://classroom.emeritus.org/courses/9054/users/229173)](https://classroom.emeritus.org/courses/9054/users/229173)

Mar 30, 2024

Marketing is a great opportunity for business to get business generated, one of the example being paper pamphlet like McDonald's , KFC does which are delivered to your mail boxes. Promoting customers to visit their store with promotions for additional food value that customers can enjoy with 50% reduction in pricing and enjoy more food. Emails also could work the same but the idea of overlooked data point could be email companies could mark them as Junk emails or end users may not look at them, and could be one of those deciding data signals.

Could be mitigated to by not looking like spams and implementing subject line in email that adds value to customers.

My 1 cent. Thank you for sharing a great discussion :)

Reply

(http

[Haitham Farag (https://classroom.emeritus.org/courses/9054/users/233864)](https://classroom.emeritus.org/courses/9054/users/233864)

Mar 31, 2024

I find applying statistics and data science in marketing to be fascinating, as it helps me grasp the implications of applying one analysis approach over another. Thanks for sharing David.

Regarding what could influence an email campaign, and using the information provided thus far, please find below some categories for your kind consideration:

1. timing of the campaign (season, beginning, end of the month etc)

2. content (use of passive, proactive language/verbs, length and ratio of text to images/pictures)

3. duration between two or more emails.

4. coinciding with a parallel marketing campaign (e.g. TV and printed, ads)

5. demographics of the target market

A main takeaway from week 1 for me, is to disaggregate data to the minutest possible level and keep running the analysis.

Edited by **Haitham Farag (https://classroom.emeritus.org/courses/9054/users/233864)** on Mar 31 at 8:36am

Reply

(http

Mariana Flores (https://classroom.emeritus.org/courses/9054/users/237198)

Mar 31, 2024

Hi David, so nice to meet you. Both the number of marketing touchpoints and campaign channel influenced sales. Were sales the main metric (KPI) to measure marketing campaign performance or success?

Curious since the marketing campaign's call to action and where it falls in the conversion funnel contributes to the number of sales. Upper funnel marketing campaigns traditionally have less sales compared to lower funnel performance marketing campaigns. Based on the call to action or business objective both could be equally successful in attaining the intended goal. How were sales attributed back to each marketing touchpoint or campaign?

Marketing campaigns or touchpoints with promotional discounts or free delivery tend to have a higher likelihood of sales just like marketing received by customers compared to prospects as there is a higher likelihood of a repeat purchase.

Such an interesting topic – thank you for sharing.

Edited by **Mariana Flores (https://classroom.emeritus.org/courses/9054/users/237198)** on Mar 31 at 9:37pm

Reply

(https:/

Lawrence Lumague (https://classroom.emeritus.org/courses/9054/users/225055)

Mar 29, 2024

Example Scenario:

A prospering financial capital is overburdened with gridlock and long travel times in the dense business districts in this town. Using vehicles has been the main source of travel since the light rail and commuter trains haven't been developed enough as to have stops in many parts of the city. Every year, more and more cars travel to these areas to shop and much of the population drive here to go to work in the office buildings within these commercial/business centers. In an attempt to reduce the influx of commuters to the business districts of a large metropolitan area, city officials have decided to implement an ordinance in which limits specific cars on heavy traffic days from Monday through Saturday.

A decision process was implemented in which Mondays, Wednesdays, and Fridays, vehicles with license plates ending with an even number cannot drive without receiving penalties and fines. On Tuesday, Thursday, and Saturday, vehicles with license plates ending in an odd number cannot drive without receiving penalties and fines.

Data Gathered by town council's data analysts prior to implementing the process: (What do we know about the inhabitants town?)

-A large population of the city is composed of mainly middle class and many upper class households due to many corporations choosing to place their headquarters in this town.

-Much of the population has the income to be able to afford shopping for goods, see shows and sporting events and do so regularly.

-The average person driving to the commercial and financial districts are composed of established families whose adults have been working for 10-15 years in which one or both adults work.

-Most of the population going to the financial and commercial business districts live in suburbs in which the popular mode of travel is driving from point ‘a’ to point ‘b’

The problem in the scenario:

The initial year of implementation of this decision was met with success by reducing vehicle congestion by 60%, though in the following years traffic was only reduced by 45% based on traffic data before this decision process was made.

The (important) question in this scenario: “Why has the effectiveness of this decision process decreased after just a relatively short period of time?”

Raw Data that was overlooked with the town’s implemented decision process:

Officials did not anticipate that both the middle-class and upper-class families will have purchased a second or even a third car, even being able to obtain license plates having one with an even number ending and the other with an odd number ending. Because of this, the effectiveness of the vehicle limitation in place was decreased, as the officials of this town had elements missing in their model that could have prevented this end result from occurring.

Having now recognized the importance of this overlooked data. Town officials can now create a new model and present it to the city council members that can help turn around the decreasing effectiveness of the town's initial model to combat traffic congestion.

Reply (1 like)

(http

Dawn Prewett (https://classroom.emeritus.org/courses/9054/users/233112)

Mar 29, 2024

This is a very interesting example and it reinforces how easy it is to miss variables that otherwise might seem obvious. For instance, the moment you explained the concept of restricting access based on odd and even plate numbers, I thought "why not just get another car". As the individual pressed with the imposing solution, my first inclination is to look for a solution to remedy the imposition. However, those working to solve the traffic problem are looking at the issue from a different focus point and therefore may become blind to the obvious loop hole. What could officials do to address this in their future decisions?

Reply (2 likes)

(http

Yossr Hammad (https://classroom.emeritus.org/courses/9054/users/229118)

Apr 1, 2024

Very interesting example. i am curious to learn what the town official could do to solve the traffic problem.

Reply

(https:/

Dawn Prewett (https://classroom.emeritus.org/courses/9054/users/233112)

Mar 29, 2024

Last year, I embarked on my first cookie season with my current Girl Scout troop. Having amassed over 12 years of data from previous cookie sales with another troop, so I was confident in my ability to predict our inventory needs, but quickly discovered new variables that I had not accounted for.

Each cookie season, scouts begin by collecting pre-orders without actual cookies in hand, a phase that lasts about a month and can serve as a barometer for customer appetite and scout interest. As the cookie chair, I then aggregate these pre-sales with historical data to forecast our initial order. This calculation is crucial as it must cover not only the pre-sales but also any personal or booth sales that occur before the "cookie cupboard" becomes accessible for replenishments. However, last year presented unexpected challenges that revealed significant gaps in my model.

Initially, I ordered 1080 boxes, a number that felt overly ambitious considering our small troop of just 4 scouts and limited number of pre-sale orders. Yet, we ran out of stock before the cookie cupboard's arrival, missing out on potential sales. This shortfall underscored two critical, overlooked factors in my predictive model: the heightened demand for Girl Scout cookies in the first normal season post-COVID and the dwindling number of Girl Scouts in our region, which inadvertently increased sales opportunities per scout. Moreover, the addition of two new scouts during the season amplified our selling power, pushing our final tally to well over 3500 boxes, which, based on data presented by parents, would have been at least 500 boxes higher had I put in a larger initial order.

As this year's cookie season began, I aimed to refine our model with these insights in mind. I used previous year's data to identify trends, determine average percentages of cookie type and percent of scout goal typically met before we started with booths to help set our initial order. Because the previous years numbers were significantly inflated due to covid, I factored in a decrease of approximately 20%. This was based on a mixture of information gathered from the East Coast sales, which occur before ours, and intuition based on experience. Due to our inventory issues the previous year, parents pushed hard for me to put in for a very large initial order, but I leaned hard on my data, which served us well as it gave us more room to maneuver and manage our inventory as the season progressed while still providing the scouts plenty of inventory for the short period between the intial order arrival and the availability of the cookie cupboard.

Despite all of this, we later discovered that we had overlooked another crucial factor: the impact of reduced sales on scout morale and their willingness to sell. This oversight was not unique to our troop but reflected a broader, area-wide trend leading to many scouts to withdraw from sales mid-season. As the weather improved and the number of scouts selling dwindled, our sales picked-up the final weekend, matching the first weekend of sales, which historically has the best sales of each season. This surge in sales enabled us to rally our troop back into action and crush our goals by over 22%.

While my model had still missed some key factors, it was strong enough to help protect our troop from costly overstock – in fact, we ended up buying two other troops out of their overstock in order to complete our final sales weekend, saving those two troops from having to purchase unsold cookies. This would not have been possible if I had started with a larger initial order and, due to having 9 different varieties of cookies, we might have ended up purchasing overstock if my variety mix had been even slightly off.

Looking ahead, I plan to adjust our sales expectations downwards and communicate more transparently with scouts to manage their expectations. I'm also exploring data-sharing with other troops to build a more comprehensive predictive model, enhancing our service unit's overall strategy. Additionally, we aim to analyze booth sales effectiveness more rigorously, moving beyond reliance on luck to informed decision-making. Last year's shift from high-cost to bargain stores as prime sales locations, along with variations in customer purchasing patterns, underscores the need for a nuanced understanding of these dynamics.

What kind of modeling changes have others had to put into place due to events like covid and what have those modeling changes looked like? What surprise impacts seemed to linger longer term than expected? I find myself curious as to how long the impact of covid will need to be included in our calculations and what we can do to safeguard against potential events like that in the future. Have you experienced other events that incurred a similar impact to covid?

Reply (1 like)

(https:/

Timothy Andrew Ramkissoon (https://classroom.emeritus.org/courses/9054/users/226697)

Mar 29, 2024

My goal was to create a data model that will predict inventory and spares levels for an operations plant and place orders for more before inventory gets low.

I configured this predictive model to use our previous Operations & Maintenance history along with analysis of data from our preservation, maintenance and storage documents and specifications to determine usage and frequency of spares needed. Also, we configured this model to track trends associated with our global facilities and sites where these assets were installed.

After implementing the model, it seems to have a fair degree of accuracy, however, the environmental conditions that the equipment were installed, to some degree, exhibited harsher conditions (higher temperatures, humidity, etc.) that were seen in our other installed locations. We are currently working on updating this model to incorporate these environmental conditions into our predictive model for a more accurate analysis of our Inventory and Spares data management strategy.

Edited by Timothy Andrew Ramkissoon (https://classroom.emeritus.org/courses/9054/users/226697) on Mar 29 at 2:47pm

Reply (1 like)

(http

Todd Engle (https://classroom.emeritus.org/courses/9054/users/228910)

Mar 30, 2024

Tim, I resonated with your predictive model, my father was a tool and die maker for many years and I grew up in machine shops. We used to have to recalibrate the machines if temperatures increased or decreased significantly from the baseline. I could see how the environment would improve the model.

I'm wondering if you also included **Operating Hours:** Equipment used for longer durations will naturally require spare parts more frequently. Demands on equipment can change increasing or decreasing operational hours. The *data point that drives demand* could be another factor.

This may not pertain to your situation, but **Operating loads** can differentiate from location to location. Equipment under heavier loads will experience wear and tear faster, requiring more frequent parts replacement. Another fine-tuning element that might help the predictive model is **Operator Training**. At the shop, newer, inexperienced machine operators tend to break machines and tools more frequently than those who are senior.

Reply (1 like)

(https:/

Mhelissa Yayalar (https://classroom.emeritus.org/courses/9054/users/233590)

Mar 29, 2024

One of the examples that comes to mind is when to launch holiday merchandizes at stores. Specifically, the launch of Christmas season decoration is appearing earlier than previous years. As a result, it becomes an advantage for businesses in the industry to launch their holiday products sooner vs later in order to gain earlier market share.

For an online retailer of Christmas décor, in order to take advantage of getting ahead in the market they would need to collect the following data:

1. Historical Sales Data-review past years sales period
2. Web traffic and online behaviors-review key words search and timing of when users started searching.

3. Online trends-such as using Google web analytics public reports to analyze past years of most popular Christmas decor
4. Historical inventory levels-determine if the qty of items sold vs unsold.

For online retailers, I think weather may not be much of a significant factor to make their predictive models accurate, but adding the economic consumer spending should be an additional data set to include.

Considering the interest rates set by government could help businesses set pricing and discounts for their products. For example, applying the same pricing structure prior to feds raising the interest rates will not be good for businesses because they could lose against competitors selling the same products for lesser price. Particularly, this year, the Fed is keeping rates unchanged, as consumer credit card and loan rates will remain high; which mean consumers will continue to tighten their spending (*Picchi, Aimee. "The Federal Reserve holds interest rates steady. Here's the impact on your money." CBSNews.com, March 20, 2024*).

Reply (1 like)

(http

Dawn Prewett (https://classroom.emeritus.org/courses/9054/users/233112)

Mar 30, 2024

You suggested that weather might not significantly influence predictive models for online retailers. Intrigued by this assertion, I delved into some research and found evidence to the contrary. Adobe, through its analytics division—which encompasses a broad range of data beyond its own—has shown that weather does indeed impact online sales. Specifically, rain can boost e-commerce spending by up to 4.4%. However, there's a threshold to this effect. When weather conditions become hazardous, this positive impact drops sharply as consumers' attention shifts away from shopping to safety concerns. This finding from Adobe marks an initial foray into understanding weather's role in online retail trends, offering valuable insights but also warranting cautious interpretation due to its novelty. (**Source: Adobe Blog on Weather Disruptions and US E-commerce in 2023 (https://business.adobe.com/blog/perspectives/weather-disruptions-will-boost-us-ecommerce-in-2023)**).

While this information is both fascinating and potentially useful for developing more accurate models, it's important to approach it judiciously, considering it represents an emerging area of analytics.

Reply (1 like)

(https:/ [Jignesh Dalal (https://classroom.emeritus.org/courses/9054/users/229173)](https://classroom.emeritus.org/courses/9054/users/229173)

Mar 30, 2024

Let's examine a scenario of an Indian Sweet store situated in a multi-culture diverse city abroad. This establishment specialized in traditional Indian sweets and hand made snacks, primarily catering to the local community and other enthusiasts of Asian cuisines. Initially, the store operates on a standard inventory model, forecasting sales based historical data, current inventory levels, and generalized seasonal trends. Despite this efforts, the store owners have notices unpredictable fluctuations in sales, particularly unforeseen surges, resulting in instances where demand surpasses supply, leading to missed revenue opportunities.

Upon a thorough analysis, a significant improvement can be suggested for the store predictive model: Integrating detailed festival calendars of Asian cultures, particular those involving consumption of sweets and snacks, such as Diwali, Eid, Pongal, and Chinese New Year. These festival witness a notable surge in demand for traditional sweets. which the store previously overlooked due to absence of specific preparation and inventory adjustments based on these cultural events.

Furthermore, it's essential to research each festival to understand the particular types of sweets and snacks preferred during these celebrations. By customizing their inventory and production plans to accommodate these preferences, the store can substantially enhance its sales and customer satisfaction during peak times and enjoy the feeling of being at home while away from home. For Example, recognizing that during Diwali there is heightened demand for sweets like Laddu and Barfi the store can increase productions of these items accordingly.

Reply (1 like)

(https:/ [Lee Lanzafame (https://classroom.emeritus.org/courses/9054/users/231975)](https://classroom.emeritus.org/courses/9054/users/231975)

Mar 30, 2024

I work for a telecommunications company.

We built a product that predicts customer experience.

We ingested survey data which easily allows us to predict if a customers having a bad experience (although it's only for 0.5% of our customer base). That initially seemed like it was a great idea but looking at the verbatims from the survey data we realised that sometimes people gave us a 7/10 because they liked the person they spoke to on the phone but they were still upset about experiencing mobile dropouts. Another limitation is that we are only allowed to survey customers once every 6 months.

We later ingested complaints data, churn data, network data, device data and many other fields. We then used SHAP values to help us determine the impact each feature had on the model. For us we knew that extrapolating this to 99% of our customer base has errors but it's still better than what we knew before the model was built.

Reply (2 likes)

(http

Jignesh Dalal (https://classroom.emeritus.org/courses/9054/users/229173)

Mar 30, 2024

Great take away from a customer service pov where customers can be felling the discomfort due to many data points but having to develop a products like CRM and collecting data can always be tedious bcoz the inputs are coming in from many data source like call centre, chat apps, websites, apps etc. Having a narrow path to understand the areas where company values lies and helping customer by getting back to them and help resolve issues for them is such a peace for customer and organization.

Personally being a consumer these days having customer feedback after every call is something that is industry standard. Great example thanks for sharing you thoughts :)

Reply

(http

Victor Flores (https://classroom.emeritus.org/courses/9054/users/197659)

Mar 31, 2024

Hi Lee - great example. Indeed, analyzing customer experience derived from complains, abandon rates, devices data and incorporating them into the data lake can yield significant conclusions which can be used to redefine where the resources of your telecom should be invested.

Reply

(https:/

Todd Engle (https://classroom.emeritus.org/courses/9054/users/228910)

Mar 30, 2024

I'm new to data analytics and, therefore do not have much real-world experience. My scenario is only theoretical and based on my own experience of having a similar user experience.

Music Recommendations for Users of a Music Streaming Service: A streaming service analyzes user listening history to recommend songs and playlists. The recommendations are hitting the mark for the most part, but one consistent complaint is too many repetitive suggestions. Including the time-of-day and day-of-the-week data could reveal users who prefer listening to the news in the morning, pop music during the afternoon, and easy listening at night. Day-of-the-week could Taylor listing habits during working days, compared to non-working days (*not everyone works M-F*).

Edited by **Todd Engle (https://classroom.emeritus.org/courses/9054/users/228910)** on Mar 30 at 6:41pm

Reply (3 likes)

(http

Jignesh Dalal (https://classroom.emeritus.org/courses/9054/users/229173)

Mar 30, 2024

Hi Todd, Thanks for sharing a great example where I have also experienced same tedious feeling of same songs being repeated. From a customer experience the model could make better suggestion like:

- Playing other songs from the same artist to see if you like the songs or one would move to next one.
- Play songs from similar artist where other users would have created a playlist and other people are playing them too.
- Timing of day is also an important aspect that could be looked away while playing from different music streaming services.
- Last but not list like suggestion like hey would you want to listen to electronic music for a change or old music from different country to explore.

There could be many suggestions from many different users but yes music streaming service model needs to be improved by making more suggestions.

Thanks for shedding some light to music. :)

Reply

(http

Mariana Flores (https://classroom.emeritus.org/courses/9054/users/237198)

Mar 31, 2024

Hi Todd, so nice to meet you. Great points – harnessing user data is key for personalized recommendations. Including recommendations based on day part, day of week, or weekend versus weekday would definitely improve recommendations and the overall user experience by more accurately aligning them to individual's preferences. To decrease the number of repetitive suggestions features like Music Genre, Artist, Album and Song could also be added. Tailoring recommendations by listeners who listened to Artist or Song Y also listened to Artist or Song X, for example. As well as having requirements in place where a song goes to the end of the queue or is only able to be recommended again after a certain time-period or number of songs may also help to decrease repetitive suggestions and create a more personalized experience while discovering new content.

Recommendation engines are fascinating - thank you for sharing.

Reply

(http

Diego Milanes (*He/Him*) (https://classroom.emeritus.org/courses/9054/users/228518)

Apr 1, 2024

Hi Todd,

This sounds like something that I need in my life :). I like to listen to the news in the morning, just as in your example, and often, I want to expand my knowledge on given topics. I'd find it helpful to add trending topics information in order to have suggestions for podcasts related to recent news.

thank you!

Reply

(http

Isabella Tockman (https://classroom.emeritus.org/courses/9054/users/207395)

Apr 6, 2024

Hi Todd,

I really like your idea. I've had the same problem with Spotify, where it feels like they just mix up the same songs in different playlists. It gets a bit boring after a while. Your thought about using the time of day and the day of the week to improve song suggestions is super smart. Adding to what you said, maybe they could also look at what kind of songs other people like us are listening to. This could make the playlists more interesting and introduce us to new songs we haven't heard before. They could also let us quickly tell them if we like their suggestions or not, so they can get better at giving us music we'll enjoy. I think your idea could really make listening to music on streaming services a lot more fun and personal. Thanks for bringing this up!

Reply

(https://

Roy Nunez (https://classroom.emeritus.org/courses/9054/users/229552)

Mar 30, 2024

We were building a model where we were trying to predict whether a customer would travel in a given month, given last 2 years of travel related purchase transactions with the goal to provide incentive recommendations and later expand into other transaction purchase categories to expand our ability to personalize recommendations. Different parameters were taken into consideration such as amounts spent, previous months travel frequency, quarter travel frequency The model predicted at a fair accuracy but were looking for features to improve accuracy.

We invited a machine learning PhD graduate to our team and he brought some important suggestions. Two features that were not initially included were 'seasonality', such as whether the customer traveled in the summer, winter, etc., and the 'amount of time since last travel', which was a piece of data that scored higher importance as a feature. These two features were included as columns and increased the model accuracy to above 93% from mid 70%s. Having other data experts join in with additional perspectives had a critical impact on our iterative model building.

Edited by Roy Nunez (https://classroom.emeritus.org/courses/9054/users/229552) on Mar 30 at 11:39pm

Reply

(http

Priscilla Annor-Gyamfi (https://classroom.emeritus.org/courses/9054/users/226376)

Apr 1, 2024

Great post Roy.

Absolutely, incorporating "seasonality" and "the amount of time since last travel" into the model design is important in addressing the problem at hand. Seasonality indeed offers valuable insights into the distinct characteristics associated with different seasons, which can greatly influence the dynamics of travel-related purchases and understanding of how customer behavior fluctuates throughout the year.

Also, considering 'the amount of time since last travel' provides crucial information about customers' travel frequency and patterns. This will help in devising effective marketing strategies and incentive recommendations. For instance, a customer who traveled recently might be less likely to plan another trip immediately, while someone who hasn't traveled in a long time might be more receptive to travel-related incentives.

Reply (1 like)

(http

Shahrod Hemassi (*He/Him*) (https://classroom.emeritus.org/courses/9054/users/224267)

Apr 1, 2024

Hi Roy. Great post!

I like the factors that you have added but think your model accuracy would improve further with some data on the customer's interests, and the type of travel that they have done in the past.

Some of this may already be addressed by the factors that you have added. For example, if the customer is traveling for winter sports, then this may be addressed somewhat by the seasonality factor. But if the customer's interest is in food & wine festivals or Formula 1 races, then these could happen at various times in the year. Factoring in the customers' interests and the events calendar for those interests may be valuable factors to improve your data model.

Reply

(http

Shahrod Hemassi (*He/Him*) (https://classroom.emeritus.org/courses/9054/users/224267)

Apr 1, 2024

Hi Roy. A couple additional suggestions:

1. Capture the customer's age and evaluate this against the frequency of travel for people in different age groups.

2. Capture where the customer lives and the frequency of travel for people living in the same area.

Reply

(http

Shahrod Hemassi (*He/Him*) (https://classroom.emeritus.org/courses/9054/users/224267)

Apr 1, 2024

Also...

3. Capture the customer's salary range and factor that against travel frequency for people at similar salary range.

Reply

(http

Roy Nunez (https://classroom.emeritus.org/courses/9054/users/229552)

Apr 2, 2024

Demographic data like age and location is not currently available to us. When we were looking at customer segmentation this is one of the first things that came up. We are working on getting this from another team even if its zip codes. Thanks!

Reply

(http

Roy Nunez (https://classroom.emeritus.org/courses/9054/users/229552)

Apr 2, 2024

Hi Sharod,

Thanks for this suggestion. Specific interests like food & wine festivals hasn't been something we have thought about. It could help our prediction models for sure. However, I am concern is too fine grained and not readily scalable. We can group similar interests though. We can use a persona-based modeling approach as well for those interests shared by a significant amount.

Reply

(http

Yossr Hammad (https://classroom.emeritus.org/courses/9054/users/229118)

Apr 1, 2024

Interesting example Roy. i agree about the importance of the length of time since last travel. When i was working at the travel organization we wanted to customize advertisements for our repeated clients and we considered this variable to group the customers based on how long since last travel.

We also grouped people by the price they paid for the last travel , and divided clients into 3 groups based on that, when we did that, the seasonality factor was not really as important since the price factor of the ticket can encourage the person to change plans and travel in different season.

These two variables helped us to run very successful advertisements and we had higher retention rate than previous years.

Reply (1 like)

(http

Roy Nunez (https://classroom.emeritus.org/courses/9054/users/229552)

Apr 2, 2024

Hi Yossr,

Thank you for the suggestions!

I agree, the seasonality factor does vary at times, as observed while we are still developing this particular model. And definitely, the length of time since last travel does hold up a strong feature importance score. We did a bit of of separation by price as some travel purchases were NYC metro card and other were flights. Will revisit while considering your thoughts. I really appreciate your feedback!

Edited by **Roy Nunez (https://classroom.emeritus.org/courses/9054/users/229552)** on Apr 2 at 11:53pm

Reply

(https:/

Mariana Flores (https://classroom.emeritus.org/courses/9054/users/237198)

Mar 31, 2024

I recently consulted for a small business owner who wanted to create a data-driven strategy to reach certain revenue goals. Understanding customers' lifetime value (CLV) is one important piece of information for companies to make business decisions as CLV is a metric that indicates the total revenue a business can expect from each customer throughout the business relationship. Additionally, organizations can identify who their most profitable segments are and where to invest by grouping customer cohorts. Grouping cohorts is traditionally done through segmentation or clustering analyses using various machine learning techniques. Parameters in these models are most commonly demographic variables while behavioral or historical purchasing data could be overlooked. Including this type of data allows the model to group cohorts into more accurate segments. Historical purchasing data like number of purchases, types of products purchased, etc. may require transformation into a usable format from raw data. This filtering, structuring, and merging of data provides a more robust set of data to draw predictions. Including historical purchasing behavior data in clustering models can help influence CLV by intelligently guiding upsell or cross-sell initiatives through data-driven recommendations and thus create a strategy to most effectively reach business objectives and goals.

Reply

(http

Victor Flores (https://classroom.emeritus.org/courses/9054/users/197659)

Mar 31, 2024

Great post Mariana! I agree on the approach of incorporating "behavioral or historical purchasing data" into a model intended to drive business recommendations to achieve higher or at least stable revenue streams for organizations. Certainly "behavioral or historical purchasing data" can be impacted by many factors which can be used to fine tune strategies to achieve milestones across times and should not be overlooked.

Reply

(https:/

Victor Flores (https://classroom.emeritus.org/courses/9054/users/197659)

Mar 31, 2024

Goal: Implement a model to predict required explosives inventory for a 3 month period to sustain production operations at a mature oilfield operated by a given national oil company (NOC)

Parameters considered by the model:

- Last 3 years production of targeted oilfield
- Aimed oil production of targeted oilfield considering existing wells and new wells to be drilled
- Competitors in the market
- Production services pricing
- Service Quality statistics (for both the company and its competitors)
- Historical inventory levels maintained by the specific district providing the services
- Inventory getting close to reach obsolescence stages

Parameter often overlooked by the model / raw data: explosives stock levels maintained by competitors. Stock levels could be impacted by many issues external to the company operations such as delays in logistic processes (clearance at origin and/or destination sties) and also by manufacturing holds at manufacturing centers. Through business intelligence practices, an estimation of the inventory stored by competitors can be obtained and incorporated into the model to allow for a more realistic planning and forecasting excercise.

Reply

(https://

Priscilla Annor-Gyamfi (https://classroom.emeritus.org/courses/9054/users/226376)

Apr 1, 2024

I agree that when developing a model tailored to a specific problem, it's possible to overlook crucial parameters that could significantly impact the relevance and precision of our solutions

A similar scenario unfolded in a project I was involved in. Our task was to devise a predictive model for a biking company to forecast bike sales and determine where to allocate investments. We factored in parameters such as revenue, expenses, and profits for each bike type, brand, state, and city over the past year. Unfortunately, we initially omitted data regarding how seasonal variations throughout the year influenced these metrics over time. Upon incorporating this information into our predictive model, we observed a marked improvement in accuracy and realism post-implementation.

Edited by Priscilla Annor-Gyamfi (https://classroom.emeritus.org/courses/9054/users/226376) on Apr 1 at 4:31am

Reply

(http

STEPHEN HUTSON (https://classroom.emeritus.org/courses/9054/users/233645)

Apr 1, 2024

Great post Pricilla!

Very interesting hearing how factoring in the seasonality data would impact the sales figures for different bikes. Perhaps another parameter that would have been interesting to incorporate would be customer reviews on the existing models to determine if there are any common themes among why certain products were reviewed better by customers, or conversely if there were common complaints about bike models that may be worth exploring additional investment by the company.

Reply

(https://

Shahrod Hemassi (*He/Him*) (https://classroom.emeritus.org/courses/9054/users/224267)

Apr 1, 2024

Scuba Dive Shop: Improving Demand Prediction to Increase Revenue & Profits

A scuba dive shop owns 2 boats and has staff for up to 4 boats. The dive shop can rent additional boats and equipment when needed. They also can bring in freelance staff when needed. At a maximum, they are able to operate 8 boats per day.

If they rent the additional boats and equipment 2 weeks in advance it costs them 50% less than the cost within 3 days of the date of rental. If they rent 1 week in advance, it costs them 25% less than the cost within 3 days of the date of rental. They are unable to cancel boat rentals. Additionally, the supply of the additional rental boats is variable (as they compete with many other dive shops) and if they wait too long, they may not be able to add additional boats. Similarly, they can hire lower cost freelance staff if they do so at least 1 week in advance or higher cost freelance staff within 1 week of the date of service. They also occasionally are unable to find enough freelance staff if they wait too long.

The dive shop is in a popular tropical tourist destination where there are a lot of hotels with people on vacation who often decide they would like to go diving with short notice (within 1-2 days). Large cruise ships also occasionally stop nearby and provide another means of demand (people wanting to dive) but they also tend to book with short notice (within 1 day).

The dive shop's goal is to maximize their profits and revenue by creating a data model that improves prediction of the demand (number of divers) so they can factor this into their advance rentals of additional boats and equipment, and their hiring of freelance staff.

The dive shop factors historical data (from previous years), the weather, the time of year, and the day of week into their data model. They have some success with this data model but still have times where they are left with empty boats that they have rented, and other times when they have to turn divers away because they do not have enough boats and equipment.

A key factor that they did not account for is the number of vacationers that were traveling to the area. They started gathering data from hotels about their upcoming occupancy rates. They also started gathering data on the planned dates for cruise ships stopping in the area and the occupancy rates of the cruise ships. Factoring in these data points into their data model, the dive shop improved the accuracy of their demand prediction and in turn, was able to improve their revenue and profits.

Question to the audience: What else could be done to further improve their data model?

Reply

(https:/

Diego Milanes (*He/Him*) (https://classroom.emeritus.org/courses/9054/users/228518)

Apr 1, 2024

At the university, the academic director implemented a data-driven model to predict the students' academic performance in their careers. This model incorporates information such as the scores from the national standardised test, the admission test and the GPA from the high school period.

This model has shown some weaknesses in describing academic performance, mainly during the first semesters, most likely due to the lack of information related to socio-economic conditions and students' adaptability to the university environment. Adding some of these features can improve the model and be used to make early detection of students with high desertion/failure risk and activate the corresponding protocols.

Reply

(http

Koffi Henri Charles Koffi (https://classroom.emeritus.org/courses/9054/users/208039)

Apr 1, 2024

hi Diego , I think an internal study need to be done to improve the performance of the students . the study need to be done a the teacher and university level too .

example how the sillabubs is made ?
how the knowledge is also share ?
what is the relationship between the student and the professor ?

here I think there is a lack of input and appropriate data to design the model that cause the system to fail.

Reply

(http

Chris Cosmas (*He/Him*) (https://classroom.emeritus.org/courses/9054/users/226607)

Apr 3, 2024

Hello Diego,

I like your case as it comes too close to home, I was a bit of a troublemaker in school and would have probably sunk the model :D

Socio-economic factors are very important, financial resources, stability at home, and social circles have a strong influence on the academic performance of students. It would be interesting to include data on external commitments such as part-time jobs, and relationship status (does the student have dependants?). Engagement also could give insightful information such as attendance, recorded misbehaviors in class, extracurricular activities, and so on. Another important factor is the workload of each course, some studies are much more cerebral and intensive than others and could affect student's performance greatly.

Reply

(https:/

STEPHEN HUTSON (https://classroom.emeritus.org/courses/9054/users/233645)

Apr 1, 2024

You are working for a city's police department and are helping them on an effort to reduce DUIs. You decide to build a model to help determine the best placement for police patrol cars to have a higher chance of catching drivers under the influence, and configure this model with past criminal cases where police have encountered drivers under the influence before. Although the initial model may be OK, in order to strengthen your outputs you incorporate a parameter that assigns a stronger weight to areas around stadiums and concert venues after events occur, because you believe there will be more drivers under the influence on the road leaving these events, who also will pose a bigger risk to the pedestrians walking out of these arenas.

Reply

(https:/

Koffi Henri Charles Koffi (https://classroom.emeritus.org/courses/9054/users/208039)

Apr 1, 2024

Colleague was approached to design a model that will predict the house cost in the Bentonville Arkansas area . To analyze and design the model 3 years of old data have been given based on those data analysis and prediction model have been designed to predict future house courses and mortgages . The model designed in the year 2020 didn't take into account the impact of the pandemic (COVID-19) that has dramatically impacted the house price , where people decide to work from home. A reconsideration by taking in consideration the pandemic and the end of the pandemic , where people are susceptible to go to office has been taken into consideration . That is to say, even if the data helps in designing the predictive model , there are future unpredicted events that can dramatically affect the performance of the model.

Reply

(https:/

Ricardo Anaya (https://classroom.emeritus.org/courses/9054/users/228915)

Apr 2, 2024

You can add more varibles to the hat examples, like What are the teams that the local team is playing against, to have those in the inventory as well (local team hats might be oversold by adversary teams).

what day of the week is it? weekday vs weekend mght also gve a different perspective,

Special events and holidays, also can contrbute to the hats selling

Reply

(http

Ricardo Anaya (https://classroom.emeritus.org/courses/9054/users/228915)

Apr 2, 2024

The more data, to correlate, the better, also ages.

you can also cross reference with people attending the stadium, to guarantee inventory and you can also create a model to price according to the demand of tickets.

Reply

(https:// Swati Sharma (https://classroom.emeritus.org/courses/9054/users/236938)

Apr 2, 2024

Problem Statement/ Use case: identifying voluntary and involuntary turnover for HR Operations

To identify voluntary and involuntary turnover , we looked at several statistics around different demographics like Turnover by gender, ethnicity, race, tenure of the employee. These metrics allowed us to understand turnover for HR operations.

However, there was one important attribute that was being overlooked, by digging deep into the data we identified another metric known as exit survey reason. Once we started to analyze our turnover metrics in regards to the exit survey reason we identified the root cause of voluntary turnover was related to compensation. This is very similar to the base ball scenario mentioned above and how Retsef illustrated the importance of digging deeper into the data.

Reply (1 like)

(https:// Isabella Tockman (https://classroom.emeritus.org/courses/9054/users/207395)

Apr 6, 2024

I work for a construction company in New York where we often deal with claims for work-related incidents. Normally, we file these claims as they come in without looking for patterns; we just report the incidents as they happen. However, our insurance company recently raised a concern. They pointed out that one of our foremen seemed to have a lot more incidents happening on his watch compared to others. They suggested we should fire him because of this. But this foreman was actually one of our best employees, and we didn't want to let him go without looking into the situation more closely.

So, we decided to take a closer look at the projects he supervised compared to those handled by other foremen. It turned out that his projects were usually three times bigger than those given to others. When we adjusted for the size of the projects, we found that the rate of incidents was actually the same as for other foremen. The only reason his numbers were higher was because he was in charge of more people and bigger projects, not because he was doing a worse job.

Reply (1 like)

(http

Gustavo Santana (https://classroom.emeritus.org/courses/9054/users/120927)

Apr 7, 2024

Thanks for sharing Isabella, really interesting how sometimes the KPIs play tricks on us. One manager I worked with used to say that data only gossip, we are the ones that need to find the truth behind it.

One example I have working in a Software House was when we were trying to find the more productive software engineers in our team, it showed that our Senior Developer spent most of his time solving bugs and not developing new features. The fact is that the bugs he was solving were not generated by him but by the junior developers. We fixed that by giving each developer their bugs to solve, making them more conscious about their work, and learning how to overcome the problems in the process.

Reply

(https:/

Gustavo Santana (https://classroom.emeritus.org/courses/9054/users/120927)

Apr 7, 2024

When working with a support team from e-commerce in Brazil, their average time for ticket conclusion was 5 days, the managers said that was too high to be true since most of the tickets were concluded in less than one day.

When we looked at the median, it was starting to get better, resulting in only 1 day, and 80% of the tickets were concluded in less than 2 days, which was the information that was presented to the directors.

The thing is the last 20% of tickets could take months to be concluded since they were not about reports by clients but intern projects from the deployment team, which took much more time to be finished. Later on, we separated them into two categories and the numbers for the support team were even better, the team was happy to celebrate their KPIs.

Reply