# **Project 1**

Mr. Adrian, the store manager at our Scotts Valley store in Minneapolis, recently concluded a pilot study in collaboration with the Workforce Management (WFM) team at Target to anonymously track and record absenteeism among his 36-member staff at the store in the past year. The absence of Team Members (TMs) makes Target’s deployment of resources and our ability to meet the needs of our guests who walk in to buy inefficient. Therefore, Adrian is keen to investigate factors that lead to TM absenteeism and further understand our need for hiring and resource deployment.



He has approached his Data Science partners hoping to help add business value by determining what causes TMs to miss extended hours and potentially implementing policies to address these issues. Additionally, Target would be able to learn the habits of TMs and schedule work hours to ensure that TMs who have a higher absentee risk are spread out among shifts to ensure a win-win scenario. He expects the team to derive findings and recommendations without any bias, and purely basis what the data shows. How well can you help him?

Data: Absenteeism\_at\_work.xlsx

**Attribute Information:**

* TM\_ID – Team Member ID; a surrogate key to identify a Team Member
* Reason for absence – the identified reason why the TM was absent from work
* Month of absence – <self-explanatory>
* Day of the week – <self-explanatory>
* Seasons – <self-explanatory>
* Transportation expense – the amount in dollars
* Distance from Residence to Work – in miles
* Service time
* Age – <self-explanatory>
* Workload Average/day
* Hit target – % of target achieved in terms of workload avg.
* Disciplinary failure (yes=1; no=0)
* Education - (high school (1), graduate (2), postgraduate (3), master and doctor (4))
* Social drinker (yes=1; no=0)
* Social smoker (yes=1; no=0)
* Pet (number of pets at home)
* Weight – <self-explanatory>
* Height – <self-explanatory>
* Body mass index – <self-explanatory>
* Absenteeism time in hours

**Group working in Project 1:** Yatam Charishma, Padmanabhuni Naga Sravya, Thondapu Lakshmi Saranya, Onteddu Poojitha Reddy

**Mentor:** Shruti GanapathySubramanian

# **Project 2**

Target, in partnership with Ingrid and Isabel, launched a five-month-long campaign, ***"Isabel Maternity by Ingrid & Isabel"***, from April to August last year with special offers on maternity clothes and a curated section on the Target.com website pertaining to the campaign. While the campaign was a success overall, the team did find potentially lost sales due to a good number of users dropping off the website without making the purchase.



The digital marketing team and creative graphics team have initiated a post-mortem study to understand how to make such campaigns more effective in the coming years and have sought help from their data science partners to carry out the analysis. They randomly sampled about twenty-four thousand user sessions and would like you to help them arrive at the root cause for the online failure rate—a success metric they abide by—and potential improvements or suggestions on what could possibly help convert more sales the next time a similar campaign is run.

Data: I&I\_2022.xlsx

**Attribute Information:**

* Year (2022)
* Month – from April (4) to August (8)
* Day – day number of the month
* Order – sequence of clicks during one session
* State – variable indicating the state within the United States where the IP address originates with the following categories:
  + 1-California
  + 2-Washington
  + 3-Massachusetts
  + 4-District of Columbia
  + 5-Virginia
  + 6-New Jersey
  + 7-Rhode Island
  + 8-New Hampshire
  + 9-Delaware
  + 10-Connecticut
  + 11-New York
  + 12-Maryland
  + 13-Illinois
  + 14-Pennsylvania
  + 15-Colorado
  + 16-Minnesota
  + 17-Oregon
  + 18-Wisconsin
  + 19-Maine
  + 20-Georgia
  + 21-Arizona
  + 22-Vermont
  + 23-Utah
  + 24-Hawaii
  + 25-Texas
  + 26-Michigan
  + 27-Ohio
  + 28-Florida
  + 29-North Carolina
  + 30-Indiana
  + 31-Kentucky
  + 32-Iowa
  + 33-Nebraska
  + 34-Missouri
  + 35-Nevada
  + 36-Kansas
  + 37-Idaho
  + 38-West Virginia
  + 39-South Carolina
  + 40-North Dakota
  + 41-Alaska
  + 42-Alabama
  + 43-Tennessee
  + 44-Montana
  + 45-Oklahoma
  + 46-New Mexico
  + 47-South Dakota
* Session ID -> variable indicating session id (a short record)
* Page 1 (Main Category) -> concerns the main product category:
  + 1 - Trousers
  + 2 - Skirts
  + 3 - Blouses
  + 4 - Jumpsuits
* Page 2 (Clothing Model) -> contains information about the code for each product (a total of 217 products)
* Colour -> colour of the product
  + 1 - Beige
  + 2 - Black
  + 3 - Blue
  + 4 - Brown
  + 5 - Burgundy
  + 6 - Gray
  + 7 - Green
  + 8 - Navy blue
  + 9 – Multi-colored
  + 10 - Olive
  + 11 - Pink
  + 12 - Red
  + 13 - Violet
  + 14 - White
* Location -> photo location on the page, the screen has been divided into six parts:
  + 1 - Top left
  + 2 - Top in the middle
  + 3 - Top right
  + 4 - Bottom left
  + 5 - Bottom in the middle
  + 6 - Bottom right
* Model Photography -> variable with two categories:
  + 1 - En face
  + 2 - Profile
* Price -> price of the product in US dollars
* Price 2 -> variable informing whether the price of a particular product is higher than the average price for the entire product category (1 if yes, 0 if no)
* Page -> page number within the Target.com website (from 1 to 5)
* Cart – whether the user added items to the cart (1 if yes, 0 if no)
* Sale – whether there was a sale recorded at the end of the session (1 if yes, 0 if no)

**Group working in Project 2:** Sreelekha Yannam, Arepalli Lakshmi Sai Teja, Meenakshi Madugula

**Mentor:** Sohini Mitra

# **Project 3**

Target operates a vast supply chain system to deliver products to guests efficiently. To improve its operational efficiency and reduce costs, the Last-Mile team reached out to the Target Facility Planning folks to put a plan together to start new Distribution Centers across the network - to optimize its inventory management process. Assume that Target only has four DCs as of today:

1. Bakersfield, CA (35.393528, - 119.043732)
2. Pueblo, CO (38.2805, -104.4672)
3. Woodbury, MN (44.979595, -93.276566)
4. Chambersberg, PN (40.028900, -77.590698)

To get the buy-in for this proposal, the Facility Planning team will need to determine the optimal number and placement of distribution centers across different regions based on guest demand patterns.



The Last-Mile team has a comprehensive dataset containing valuable information about guest orders, geographical locations, and other relevant attributes and are seeking help from the in-house Data Science team to analyze the provided dataset and employ techniques to make informed decisions regarding the optimal placement of distribution centers. The goal is to identify regions with similar guest demand patterns and determine the appropriate number of distribution centers needed to fulfill guest orders effectively. You, as part of the team who is driving this solution are therefore required to present you findings, including the recommended number of distribution centers and their proposed locations (don’t need the exact lat-long precision, just the state would do) based on the identified demand patterns.

To get you started, the Last-Mile team has compiled some tasks to help you build the solution:

* Identifying regions with high guest concentrations and frequent order placements to prioritize distribution center placement in those areas.
* Analyze the average order values and product preferences of guests in different regions to optimize inventory management and stock allocation.
* Consider the returns rate as a factor to minimize return shipping distances and associated costs when deciding on distribution center locations.
* Incorporating the geographic locations of guests to ensure efficient and timely product deliveries.

Data: DC\_Planning.csv

**Attribute Information:**

* Guest ID: The unique identifier for each guest in the dataset. It follows the format 'GXXXXX', where 'XXXXX' represents a guest identification number. Sample values: 'G12345', 'G67890', 'G54321'.
* Geographic Location: The geographic location of each guest, represented as latitude and longitude coordinates. Sample values: (37.1234, -122.5678), (39.5678, -118.9876), (35.9876, -120.3456).
* Order Frequency: The frequency of guest orders, indicating how often a guest makes a purchase. This value is a positive decimal number. Sample values: 8.93, 12.45, 9.17.
* Average Order Value: The average value of guest orders, representing the average amount spent per order. This value is in monetary denomination (e.g., dollars, euros). Sample values: 45.67, 89.23, 52.10.
* Electronics: Binary variable indicating the guest's preference for electronics products. 1 represents preference, and 0 represents no preference. Sample values: 1, 0, 1.
* Clothing: Binary variable indicating the guest's preference for clothing products. 1 represents preference, and 0 represents no preference. Sample values: 0, 1, 1.
* Home Appliances: Binary variable indicating the guest's preference for home appliances products. 1 represents preference, and 0 represents no preference. Sample values: 1, 0, 0.
* Books: Binary variable indicating the guest's preference for books. 1 represents preference, and 0 represents no preference. Sample values: 0, 1, 0.
* Beauty: Binary variable indicating the guest's preference for beauty products. 1 represents preference, and 0 represents no preference. Sample values: 1, 1, 0.
* Returns Rate: The rate of returns for each guest, representing the proportion of items returned compared to the total number of items purchased. This value ranges between 0 and 1. Sample values: 0.15, 0.03, 0.10.
* Guest Segmentation: The segmentation or categorization of guests based on certain criteria or characteristics. Possible values are "Repeat," "Reactivated," "Lapsed," and "New". Sample values: "Repeat," "Lapsed," "New".
* Shipping Preferences: The guest's preferred shipping method for deliveries. Possible values are "Standard," "Express," and "Same-day". Sample values: "Express," "Standard," "Same-day".
* Promotional Response: The response level of guests to different types of promotions or marketing campaigns. Possible values are "Low," "Medium," and "High". Sample values: "Medium," "Low," "High".

**Group working in Project 3:** Gopisetti Sri Aishwarya, Asritha Bhimavarapu, B Sita Swapnika, Bhavitha Sri Kopparapu, Sreevidya Bagareddgari

**Mentor:** Aviral Maheshwari

# **Project 4**

Problem Statement:

One of the most important aspects of running a new business is to innovate to stay ahead of the competition. Every year target invests millions of dollars on different marketing campaigns to boost awareness , increase brand loyalty, enhance reach and get more sales from target’s customers. Target’s Marketing team would like to understand the impact of various features on the clicks acquired each day for a campaign. You are given a set of 984 campaigns from the past year along with some of the features.



The Data Includes the following columns –

ad\_grp\_id – unique Id of the campaign

category\_name – category of the products for which the campaign is running for.

report\_date – date of the data

bids – No. of bids

impressions – No. of views for the advertisement

cost – Total Spend to get the Number of views

clicks – Total Clicks on the advertisement

bid\_cost – Total Amount of Bid Cost in dollars

The team would like you to create combination of features from the above data available and tell the impact of change of those features on the clicks that an advertisement gets. Also Comment on how the nature of the impact changes as you move from a day to week to a month time frame.

You should be able to understand all the terms and learn more about the ad tech industry here.

Dataset: Adtech.csv

**Group working in Project 4:** Amrutha Nrusimhadri, Kasula Spandana, K. Sai Tejaswi, Magatala Nimitha Reddy, Kaandru Nithya

**Mentor:** Vijayendra Grampurohit