**DSSP Project 1**

**Unlocking Insights from Telecom Usage Data:**

**Predicting Customer Value and Behavior**

**Introduction:**

In the highly competitive telecommunications industry, understanding customer behavior and value is crucial for optimizing marketing strategies, improving customer retention, and maximizing profitability. Your challenge is to analyze rich, multi-source telecom datasets containing detailed customer demographics, geographic information, device characteristics, and usage statistics to unlock value for AT&T and their customers.

The available data consist of 6 csv files covering different aspects of customers and their interactions. Your goal is to clean and wrangle these datasets, formulate meaningful questions, and apply statistical and machine learning techniques to extract actionable insights that could help the telecom company better understand and serve their customers. Remember the two fundamental business questions:

* + What specific action(s) do we need to take?
  + What is the value? How much money will this/these actions save or make? What is the impact on revenue, customer experience, etc?

**Data Description**

The project data are distributed across **six files**, each providing a unique perspective on the customers and their usage patterns:

1. **Area Code Data**: Geographic area codes linked to customer locations.
2. **Prizm Customer Data**: Customer segmentation data based on socio-demographic profiles.
3. **Telco Customer Geographic Data**: Customer Geographic information.
4. **Telco Phone Data**: Device-related information such as handset type, age, and features.
5. **Telco Usage Data**: Detailed monthly usage metrics, revenue figures, and service quality indicators.
6. **Telco Data Dictionary**: Definitions and explanations of all variables.

**Project Instructions**

**Explore & Understand the Data**  
Thoroughly investigate the six data files to understand their contents and relationships. Consult the Telco Data Dictionary for variable definitions and units. Important note: you are not expected to use all the data in the files provided. In fact, it is very much expected that you will use only a subset in your focused analysis. For instance, there are many potential response variables including, total revenue, total minutes used many others. A very successful project may only look at one of these.

**Identify One or More Research Questions**  
This project is **open-ended**: you may investigate a single inference or multiple questions. Possible themes include, but are not certainly not limited to:

* + Predicting customer revenue or lifetime value.
  + Segmenting customers based on usage and demographic profiles.
  + Evaluating the impact of device type or geographic region on usage patterns.
  + Assessing service quality effects (e.g., dropped calls) on customer behavior or revenue.

**Select Your Response Variable(s)**  
The Telco Usage Data contains several potential response variables, such as various revenue measures, minutes of use, or service quality indicators. Choose the response(s) that best fit your research questions.

**Be Mindful of Variable Dependencies**  
For example, total revenue can often be calculated from other revenue-related columns. Think carefully about what information is realistically available at prediction time and avoid data leakage.

**Apply Appropriate Analytical Methods**  
Use statistical inference, regression, classification, clustering, or other suitable methods. Justify your choices and validate your models rigorously.

**Tell a Story**  
Build a clear, insightful narrative around your analysis. Provide business context and actionable recommendations where possible.

**Communication is Key**  
Communication within your group is very important. Usually, when groups run into problems, it is because of poor communication. Please respond promptly to emails and other outreach efforts from your teammates and the instructor. Most groups work very well; when issues arise, they are almost always due to communication breakdowns. Stay connected and proactive!

**Deliverables (Submit to Canvas)**

1. **IPython Notebook**
   * Well-commented code and narrative explanations.
   * Data cleaning and preprocessing steps.
   * Exploratory data analysis and visualization.
   * Modeling and inference results.
   * Interpretation and discussion of findings.
2. **Professional Presentation**
   * Summarize your work in a clear, concise, and engaging slide deck (PowerPoint, Google Slides, or similar).
   * Highlight key insights, visualizations, and business implications.
   * Ensure the presentation is polished and suitable for a business audience.
   * Each group will have 15 minutes to present with 5 to 10 minutes of questions.

**Timeline & Support**

**Meetings During Week 3**  
I will check in with each group at some point during Unit 3 to answer any questions and provide support. It is important to clearly identify your question(s) of interest and to make sure everyone in your group is on the same page and coordinated.

**First Class of Unit 4 Check-in: Tuesday, June 10**  
I will meet with each group individually on Zoom to discuss progress, ideas and next steps.

**Live Practice: Thursday, June 12**  
I will be in town on Thursday, June 12 to meet with each group in person and provide feedback on your presentations. The more live practice the better! ☺

**Project 1 Presentations! Friday, June 13, 1pm to 4pm CST in Frances Moody 142**  
Mark Austin as well as senior level data scientists will be present during your presentation.

**Office Hours / Meetings**  
I am available by appointment to discuss group questions and provide guidance. Please reach out early if you would like to discuss your ideas!

**Things to Keep in Mind!**

Depth and rigor of data exploration.

Clarity and creativity in research questions.

Appropriateness and correctness of analytical methods.

Quality and clarity of communication in notebook and presentation.

Practical relevance and insightfulness of conclusions.

Clean slide deck.

**Groups for Project 1!**

* Group 1: Bijay, Blake, Maggie
* Group 2: Margarat, Ria, Robert
* Group 3: Amanda, Marione, Tommy