STORE
SIMULATOR

HTTPS://GITHUB.COM/STEPHANIE-SCHEFER/AUTOMATION_FINAL_PROJECT

AGENDA

01 PROJECT MOTIVATION

02 STORE DATA OVERVIEW

03 STRATEGY

04 DEMONSTRATION

05 CHALLENGES

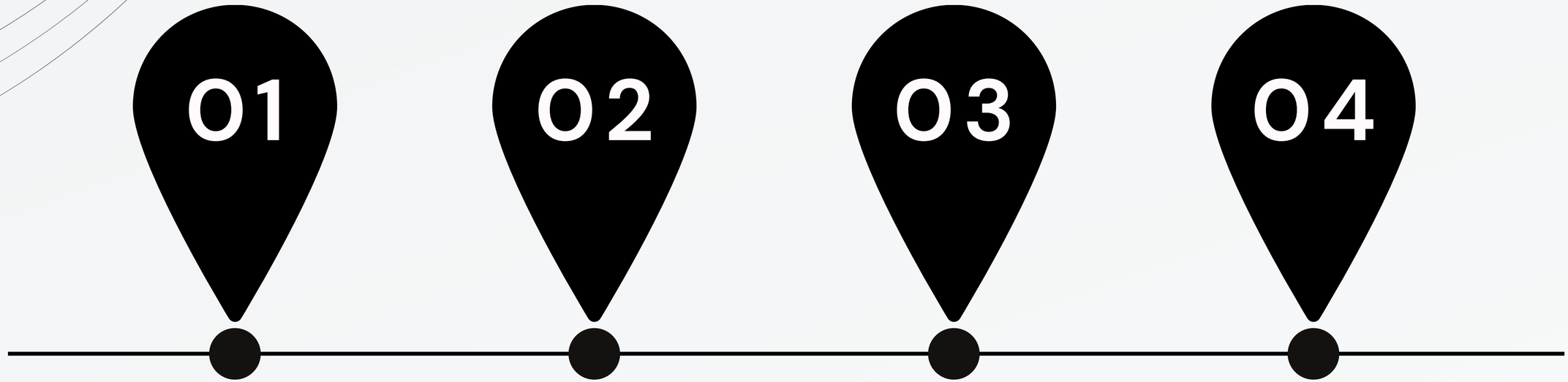
06 FUTURE ITERATIONS

PROJECT MOTIVATION

- **My background:** Entrepreneur and a Business Analytics and Data Science Double Major
- Analyze sales data to gain insights into consumer behaviors
- Improve project workflow and ability to work with classes in python
- **“Girl Math”:** applying funny logic to rationalize why an individual is getting a good deal as far as time, money, or convenience
  - ex. buying an expensive accessory and justifying the purchase with a per day dollar amount after frequent use for a duration of time.
- **Problem:** lack of publically available transaction data

STORE DATA OVERVIEW

DATA SIMULATION



CUSTOMERS

- ID
- Name
- Age
- Gender
- Location
- **Buyer Habit** (routine/impulse)
- **Buyer Spending** (heavy/moderate/light)

PRODUCTS

- ID
- Category (Office Supplies, Accessories, Beauty, Personal Care, Clothing, Groceries, Home, Electronics, Shoes)
- Price
- **On Sale** (True/False)
- **On Display** (True/False)

TRANSACTIONS

- ID
- Customer ID
- Purchase Time
- Total Transaction Cost

TRANSACTION DETAILS

- ID
- Product ID

STRATEGY

OVERALL PROCESS

- Import Data
  - Names and Zips
- Generate Customer and Products
  - Save Output
- Sales Generator
  - Transactions
  - Transactions Details
  - Save Output

SIMULATION

- Create a Database For The Store Data
- Transform Tables Into a Usable Table (Stats by Customer)
  - Purchase Frequency
  - Total Spent
  - Average Spent
  - Average Variance in Cost
  - Category Distribution
  - Total Products Purchased

FEATURE ENGINEERING

- Customer Segmentation Analysis
  - Identify impulse purchasers
- Test Various Models
  - Random Forest
  - Logistic Regression
  - KNN
  - Gaussian Naive Bayes
  - Support Vector Machines
- Evaluate model accuracy through training and testing

MACHINE LEARNING

DEMONSTRATION

CHALLENGES

Finding Data

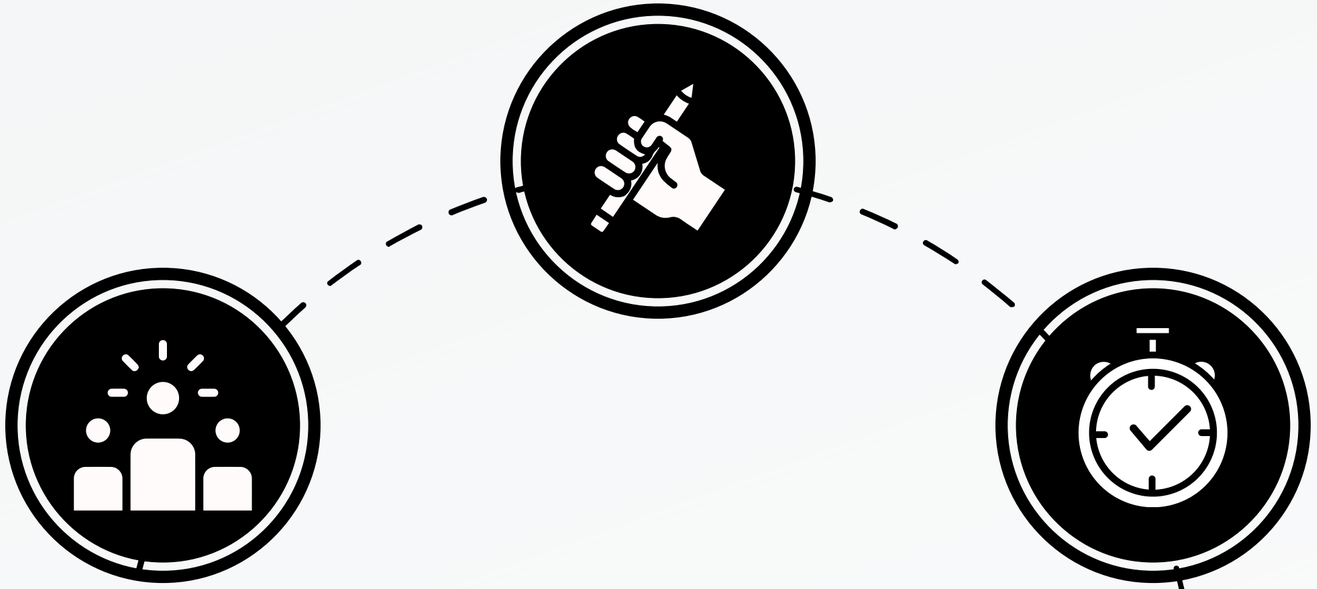
- Altered Project Plan
- More time spent on the simulation rather than the analysis compared to my prior expectations

Product Pricing Matches Total Spent

- Sales volume must increase over time
- Allow economic conditions and time intervals to influence the number of transactions per day
- Transaction price must correspond to individual consumer buying trends (products and total cost)

Run Times

- Running the simulation for a year's worth of data takes over 6 hours to run locally



FUTURE ITERATIONS

- Generate new sales
- Update database
- Re-segment customers

MONTHLY WORKFLOW

- Automate the process currently outlined in the testing_code.ipynb file

MACHINE LEARNING AUTOMATED PROCESS

- Streamlit Webpage
  - Set Parameters
    - Help Eliminate Potential Input Errors
  - Click Run
  - See Results

VISUALIZATIONS