## Scenario

<https://www.pexels.com/photo/person-showing-bedroom-interior-237371/>

A picture containing indoor, floor, bed, wall

Description automatically generated

You are the lead marketing data scientist for Bedding Bathing & Yonder (BBY). BBY is an American chain of domestic merchandise retail stores with an online presence. The chain primarily operates stores throughout the United States. BBY offers a membership and loyalty program designed to give customers easy access to great benefits and discounts through email, and physical direct mail coupons tailored to the household shopping habits. Loyalty customers have shared some basic information about their household and as a result the company’s marketing department has excellent data assets for these customers.

Senior management has asked you make an in-depth exploratory analysis of the customer base. The goal of the management team is to create micro-segments and personas for the loyalty program. Keep in mind, interesting data correlations may not be beneficial in a marketing context. For example, identifying 5 customers with very specific attributes may be interesting but hardly a segment worth attracting. Further, this data has been randomized due to privacy concerns so some attributes may not be useful.

**You are asked to examine the data, clean it, use supplemental data to enrich the data then identify 4 or more interesting insights from the user data. All relevant cleaning, enriching and EDA steps along with the 4+ insightful data nuances should be organized into a presentation. You will present to the head of marketing who is looking for an “ah –ha” persona or previously unknown data relationship. As the head of marketing, relevant information is consumed visually instead of in table form. Thus, your presentation should include visualizations when appropriate. Your submission will include code, bulleted written document and PowerPoint slides with narration.**

## Data

Source: The data has been synthesized using existing proprietary household data sets from various third parties obtained for a single US community. Some variables have been randomized, others have been anonymized, and further de-identification is performed on the geolocation attributes (lat/lon these have been completely manufactured). Thus, the data is likely not able to be reconstructed in its initial proprietary form but is still representative in some regards.

## Data Tables

|  |  |  |  |
| --- | --- | --- | --- |
| File Name | Rows | Columns | Description |
| consumerData\_training15K\_studentVersion.csv | 10000 | 26 | Consumer purchasing habits |
| DonationsData\_training15K\_studentVersion.csv | 10000 | 15 | Household donation history |
| inHouseData\_training15K\_studentVersion.csv | 10000 | 20 | Data BBY has as part of membership |
| magazineData\_training15K\_studentVersion.csv | 10000 | 10 | Household magazine subscription history |
| politicalData\_training15K\_studentVersion.csv | 10000 | 13 | Household political leanings |

All data tables have a unique identifier, "tmpID" which can be used to join the data. If joined properly, the training data will have 80 variables described below. It may not be the case all variables are useful or even ethical to use for insights, EDA and/or persona building.

## Data Dictionary

|  |  |  |
| --- | --- | --- |
| **Variable Name** | **Example** | **Definition** |
| tmpID | 5126 | Unique Household Identifier |
| ResidenceHHGenderDescription | Female Only Household | Head of household gender |
| EthnicDescription | Chinese | ethnic description |
| BroadEthnicGroupings | German | secondary ethnic description |
| PresenceOfChildrenCode | Known Data | If known, children purhcases indicate a child in home |
| ISPSA | 8 | Index of Social Position for Small Areas (ISPSA) Index of Social Position for Small Areas (ISPSA) an exclusive lifestyle classification system providing household characterization by social class affiliation. ISPSA is a score weighted on several census factors related to education and occupation. |
| HomeOwnerRenter | Likely Homeowner | Home dwelling type |
| MosaicZ4 | Bohemian Groove | Persona Descriptions from Experian  - i.e. https://assets.cengage.com/gale/help/dnow/Mosaic/MosaicTypeK40\_DescPortrait.pdf |
| MedianEducationYears | 13 | Number of formal education in years |
| NetWorth | $50000-99999 | Binned, net worth of household |
| Investor | Yes | Indication of an investor in the household |
| BusinessOwner | Yes | Indication of a registered business owner in the household |
| Education | Grad Degree - Likely | Modeled highest formal education attained |
| OccupationIndustry | Unknown | Head of household professional industry |
| HorseOwner | Yes | Indication of a equestrian rider and/or owned horse |
| CatOwner | Yes | Indication of a cat(s) in household |
| DogOwner | Yes | Indication of a dog(s) in household |
| OtherPetOwner | Yes | Indication of a non cat/dog/horse pet in household |
| HomeOffice | Yes | Indication of a home work office |
| BookBuyerInHome | 3 book purchases in home | Household book purchases from book-scan |
| UpscaleBuyerInHome | 1 upscale merchandise purchase | Household luxury goods pruchase indicator |
| BuyerofAntiquesinHousehold |  | Household antique purchase indicator |
| BuyerofArtinHousehold | Yes | Household fine art purchase indicator |
| GeneralCollectorinHousehold | Yes | Household collector purchases indicator i.e. cards, stamps |
| BooksAudioReadinginHousehold | Yes | Audio book purchase indicator |
| ComputerOwnerInHome | Yes | Presence of a home computer |
| ReligiousContributorInHome | 1 religious contribution in home | Donation to a religious charity indicator |
| PoliticalContributerInHome | 1 political contribution in home | Donation to a political nonprofit, candidate or elected official |
| DonatesEnvironmentCauseInHome | Unknown | Donation to environmentally focused non-profit |
| DonatesToCharityInHome | Unknown | Donation to a non-profit of any kind in household |
| DonatestoAnimalWelfare | Yes | Donation to ethical treatment of animals nonprofit indicator |
| DonatestoArtsandCulture | Yes | Donation to arts, performing, and culture related nonprofit organizations |
| DonatestoChildrensCauses | Yes | Donation to child advocacy and welfare non profit organizations |
| DonatestoHealthcare | Yes | Donation to healthcare and wellbeing organiations |
| DonatestoInternationalAidCauses | Yes | Donation to international, non-US causes i.e. relief agencies |
| DonatestoVeteransCauses | Yes | Donation to conflict and war veteran support organizations |
| DonatestoHealthcare1 | Yes | Donation to healthcare and wellbeing organiations |
| DonatestoInternationalAidCauses1 | Yes | Donation to international, non-US causes i.e. relief agencies |
| DonatestoWildlifePreservation | Yes | Donation to environmentally focused non-profit |
| DonatestoLocalCommunity | Yes | Donation to county, zip and community organizations |
| FirstName | Isiah | Fictitious first name for the head of the household |
| LastName | Bechtelar | Fictitious last name for the head of the household |
| Gender | M | BBY loyalty member gender |
| Age | 49.55844156 | BBY loyalty member age |
| TelephonesFullPhone | (539)-471-5789 | Fictitious telephone number (US) |
| lat | 48.29635 | latitude |
| lon | -95.54055 | longitude |
| county | Beltrami County | County (administrative division of a state, providing certain local governmental services) name |
| city | Grygla | City (administrative municipality) name |
| state | Minnesota | US state name |
| fips | 27007 | Federal Information Processing System (FIPS) Codes for States and Counties.  FIPS codes are numbers which uniquely identify geographic areas. |
| stateFips | 27 | FIPS state code |
| HomePurchasePrice | $243,000 | County register record for home price in last sale |
| LandValue | $11,000 | County register land deed assessed value |
| DwellingUnitSize | 1-Single Family Dwelling | Occupancy Type |
| storeVisitFrequency | 7 | Number of times loyalty member has been in store in last 12 months |
| PropertyType | Residential | Household property type |
| EstHomeValue | $187,500 | Estimated home value including land according to third party, non-county register model |
| yHat | 123.0180871 | Average household spend with BBY in USD. Some noise has been added to the amounts for further anonymization |
| FamilyMagazineInHome | 2 family-oriented magazine purchases | Magazine subscription related to family relations in house |
| FemaleOrientedMagazineInHome | 1 female magazine purchase | Magazine subscription with female target in house |
| ReligiousMagazineInHome | 1 religious magazine purchase | Magazine subscription on religious topic in house |
| GardeningMagazineInHome | 2 gardening magazine purchases | Magazine subscription covering gardening, and landscaping in house |
| CulinaryInterestMagazineInHome | 1 culinary magazine purchase | Magazine subscription focused on fine dining, diet, cooking or baking |
| HealthFitnessMagazineInHome | 4 health and fitness magazine purchases | Magazine subscription focused on health and wellbeing, working out, yoga or other fitness/active/sport related activities |
| DoItYourselfMagazineInHome | 3 Do-It-Yourself magazine purchases | Magazine subscription focused on home improvement, woodworking, or other self initiated projects |
| FinancialMagazineInHome | 1 financial magazine purchase | Magazine subscription focused on retirement, investment opportunities or retail equity |
| InterestinCurrentAffairsPoliticsInHousehold | Yes | Magazine subscription focused on newworthy events, local or global in nature |
| PartiesDescription | Republican | Registered political party affiliation |
| ReligionsDescription | Protestant | Likely religious denomination |
| LikelyUnionMember | Yes | Likely union roster appearance |
| GunOwner | Yes | Registered gun owner |
| Veteran | Unknown | Likely foreign war veteran or member of the armed service |
| supportsAffordableCareAct | Support | Political support for nationalized US healthcare |
| supportsGayMarriage | Support | Political support for LGBTQ+ rights |
| supportsGunControl | Oppose | Political support for gun ownership |
| supportsTaxesRaise | Oppose | Political support for raising taxes among wealthy households |
| overallsocialviews | Conservative | Overall classification of political leanings in the head of the household |
| DonatestoConservativeCauses | Yes | Donates to traditionally conservative cause nonprofit organizations |
| DonatestoLiberalCauses | Yes | Donates to traditionally liberal cause nonprofit organizations |

*Further data enrichment is possible using FIPS code but is not required for the case. The US Census publishes data using the FIPS code and some online data sets exists for community level attributes. However, the FIPs in this data set are incomplete and do not get to the block level. As a result, additional geospatial work is not required to complete the case. Exclusive to the case, for those that are interested, some starting links are shared below.*

<https://data.world/datasets/fips>

<https://www.huduser.gov/portal/datasets/usps_crosswalk.html>

## FIPS Code Explanation

A picture containing diagram

Description automatically generated

*Source:* [*https://www.policymap.com/2012/08/tips-on-fips-a-quick-guide-to-geographic-place-codes-part-iii/*](https://www.policymap.com/2012/08/tips-on-fips-a-quick-guide-to-geographic-place-codes-part-iii/)

## Course Scripting Supplemental

You will receive an initial script with code examples to get you started since this is the first case of the course.

## The Submission

* The submission will include business analyst slides covering the problem, data and 4 insights. Without a presentation, the “organization” section of the rubric will be 0. Exceptional submissions are well ordered and provide a coherent narrative covering all 4 insights.
* The submission will include a written supplemental representing the 4 insights identified and described in the business presentation. The written portion can be 3-5 sentences for each insight in a bulleted list format. Exceptional submissions include statistics from external credible sources that support the identified personas or insights. For example, “…focus on pet owners over age XX because [some org/research] says this segment will grow YY over the next 5 years…”. Without a written supplemental that coincide with the narration and supported by code the “written supplemental” section will be 0.
* The submission will include either a recorded screen narration of the business presentation, a text file with a URL to a recording (like youtube video) or audio that is embedded into the slide deck. Tone, volume, cadence, use of filler words and pronunciation will be accounted for in this section. No points will be deducted based on English proficiency (ie ESL) but technical descriptions that are incorrect will be detrimental. Failure to submit a narration, the “delivery” section of the rubric will be 0.
* An R script covering all data munging, modeling (if applicable), evaluation (if applicable) and visualization construction used to create the presentation artifacts (you do not need to use R to construct the slides but it is possible) and come to the case outcomes. Your code must use the following R functions at least once throughout your code, group\_by, aggregate & subset. Make sure to that your code contains ample comments. Failure to turn in an R script will result in a “Documentation” score of 0.

## Criteria for Success

The presentation will be evaluated on an equal weighted scale with the following criteria. For example, 20 points per each category [depends on the individual course weighting found in Canvas]

* **Organization** – Was the presentation well organized?
* **Delivery** – Was the content delivered clearly and persuasively with the audience in mind?
* **Code Documentation** – Was the data mined to support the conclusion?
* **Written Supplemental** – Are the bullets clear and supported in narration and code?
* **Data Mining Proces**s – Overall, as a complete portfolio of work, is the topic interesting, organized, researched, supported and delivered effectively?