

Введение

Файл «Bike Buyers.xlsx» – это база данных, содержащая информацию о покупателях велосипедов.

Цель данного проекта – произвести очистку данных (Data Cleaning) и анализ данных (Data Analysis) с визуализацией данных (Data Visualization).

Метаданные

Набор данных представляет собой файл в формате Microsoft Excel и представлен следующими столбцами:

- ID (General) – идентификационный номер покупателя
- Marital Status (General) – семейный статус
- Gender (General) – пол
- Income (доход) – доход
- Children (General) – количество детей
- Education (General) - образование
- Occupation (General) – занятость
- Home Owner (General) – владелец дома
- Cars (General) – количество машин
- Commute Distance (General) – дистанция до работы
- Region (General) – Регион
- Age (General) – возраст
- Age Brackets (General) – возрастная категория
- Purchased Bike (General) – приобрел велосипед

Файл содержит 1027 строк и 13 столбцов. Все операции были выполнены в Microsoft Excel. Все файлы доступны по ссылке: [enupilov/BikeBuyers \(github.com\)](https://github.com/enupilov/BikeBuyers).

Очистка данных (Data Cleaning)

При очистке данных были выполнены следующие шаги:

1. Созданы листы Working Sheets, Pivot Table, Dashboard;
2. Удалено 26 дубликатов строк;
3. Заменены буквенные обозначения слов в столбцах Marital Status и Gender на полные слова для лучшего понимания;
4. Стандартизированы столбцы по типу данных;
5. Добавлен столбец Age Brackets (Возрастная категория) для последующей визуализации;

Анализ данных (Data Analysis)

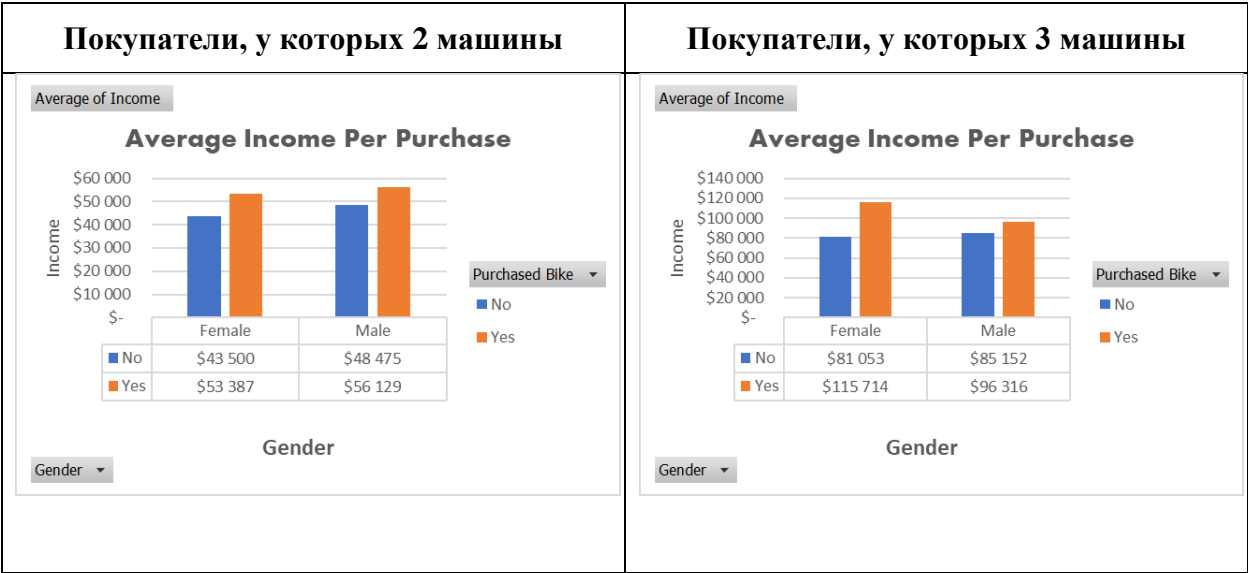
При анализе данных были выполнены следующие визуализации:

1. Созданы сводные таблицы с графиками;
2. Создан Дэшборд.

Средняя заработная плата покупателей

В разрезе заработной платы было выявлено, что мужчины зарабатывают больше женщин. Для обоих полов характерно приобретение велосипедов. Примечательно, что в регионе Тихого океана (Pacific) одинокие женщины зарабатывают больше, чем мужчины и менее предпочтительны к приобретению велосипеда. Похожая ситуация наблюдается и в Европе.

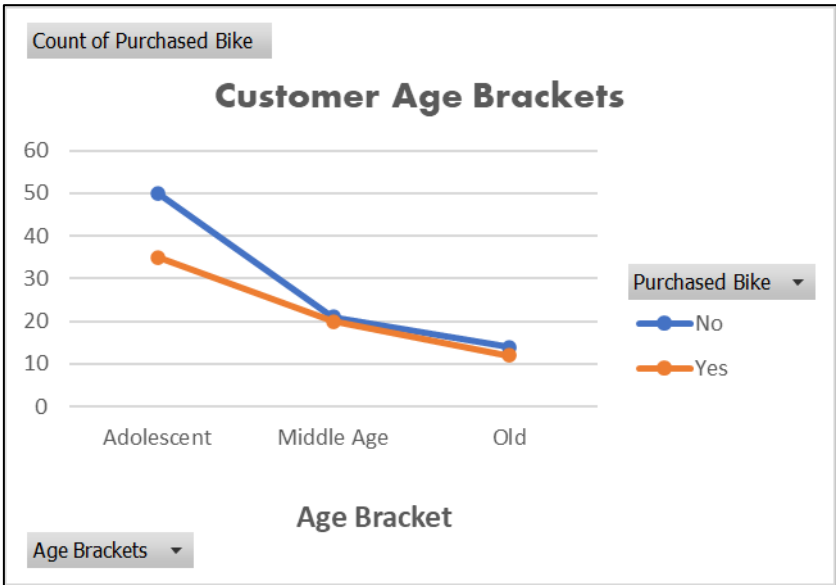
Также хотелось бы отметить, что доход резко возрастает при наличии у покупателя больше 2 машин.



Возрастная категория покупателя

В разрезе возраста покупателей отмечается, что с возрастом люди менее предпочтительны к приобретению велосипеда. Но одинокие люди все же более предпочтительны к покупке, чем семейные.

Все же наличие машины имеет значение если у одинокого человека как минимум 2 машины он не заинтересован покупать велосипед.



Дистанция до работы

В разрезе пути до работы можно отметить, что люди в основном работают в миле от дома. Данная дистанция наиболее предпочтительна для покупки велосипеда.



В разрезе образования примечательно, что люди с неполным высшим образованием предпочитают не кататься на велосипеде. За исключением тихоокеанского региона. Там люди предпочитают велосипед. Но он становится менее предпочтителен, когда расстояние превышает 5 миль.



Заключение

Таким образом, можно отметить, что мужчины более предпочтительны к приобретению велосипеда. С возрастом желание купить велосипед уменьшается. Также использование велосипеда на дальние расстояния вносит определенные неудобства по сравнению с использованием автомобиля, что несомненно влияет на решение приобрести велосипед или нет.