

## Обработка результатов опроса

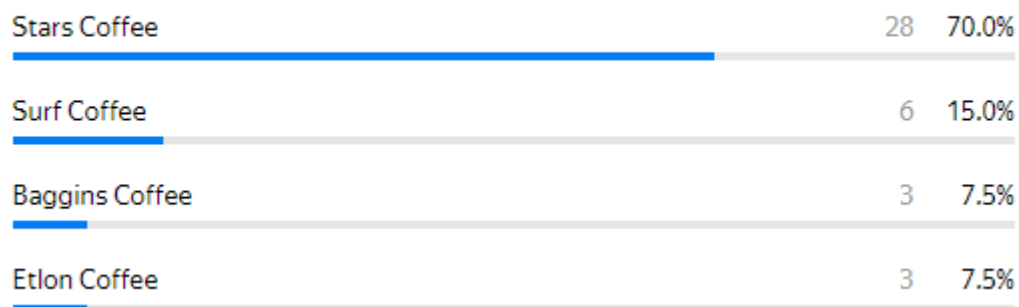
Среди кофеманов Санкт-Петербурга был проведен опрос на 4 вопроса, для выявления следующих вопросов:

- Где наибольшее разнообразие кофейных напитков
- Где кофе наиболее крепкий
- В какой кофейне наибольшее разнообразие альтернативного молока
- Где цена соответствует качеству

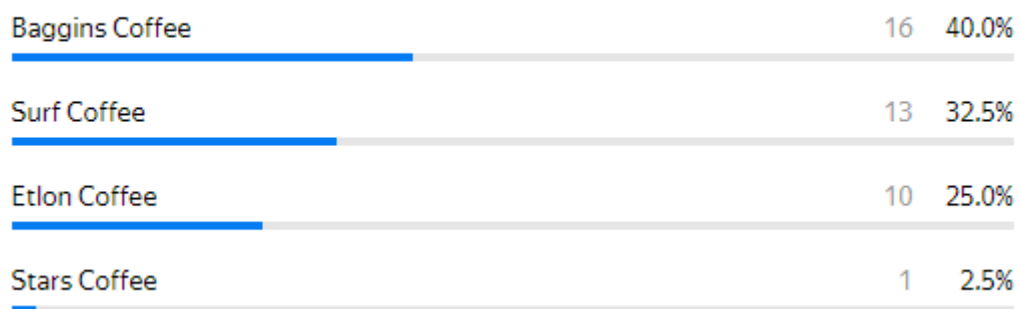
Было представлено 4 варианта ответа: Baggins Coffee, Stars Coffee, Etlon Coffee, Surf Coffee.

В результате опроса, мы получили 40 ответов:

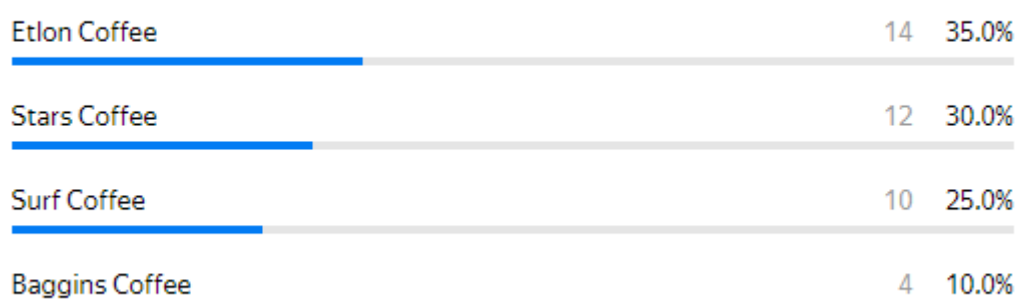
Где наибольшее разнообразие кофейных напитков ?



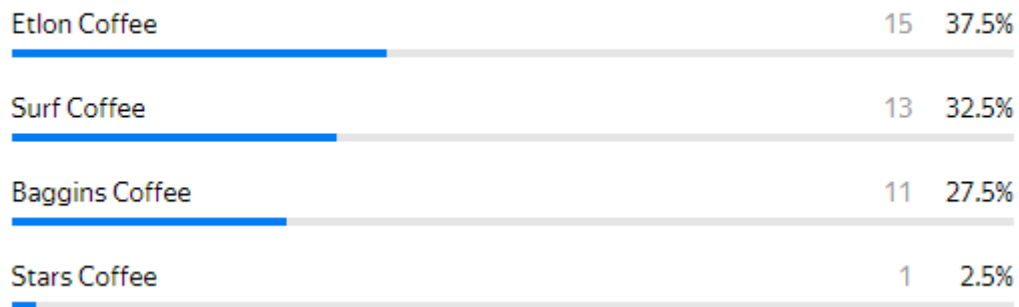
Где кофе наиболее крепкий ?



В какой кофейне наибольшее разнообразие альтернативного молока ?



Где цена соответствует качеству?



Для обработки результатов используем язык программирования Python, а точнее его библиотеки: matplotlib, pandas, scipy, itertools.

Собранные данные опроса сохранили в формате json и отправили в VS code для чтения и обработки:

```
json_file_path = 'data.json'

# Чтение данных из JSON-файла
with open(json_file_path, 'r', encoding='utf-8') as file:
    json_data = json.load(file)
```

Для более удобной работы, преобразовываем данные в DataFrame:

```
# Преобразование в DataFrame
columns = [item[0] for item in json_data[0]]
data = [[item[1] for item in row] for row in json_data[1:]]
df = pd.DataFrame(data, columns=columns)
```

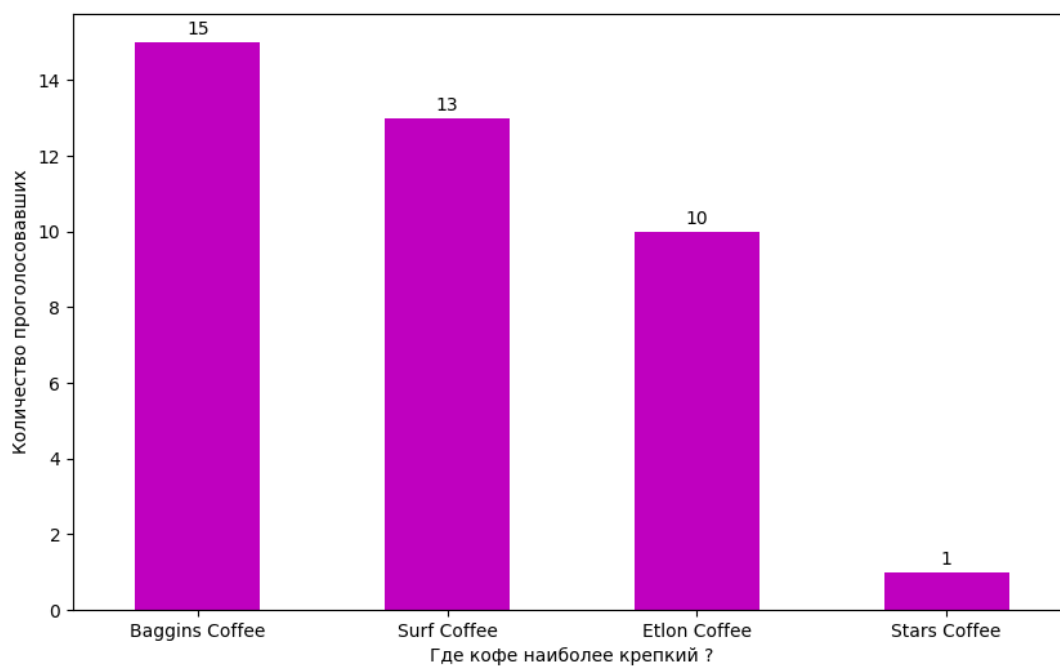
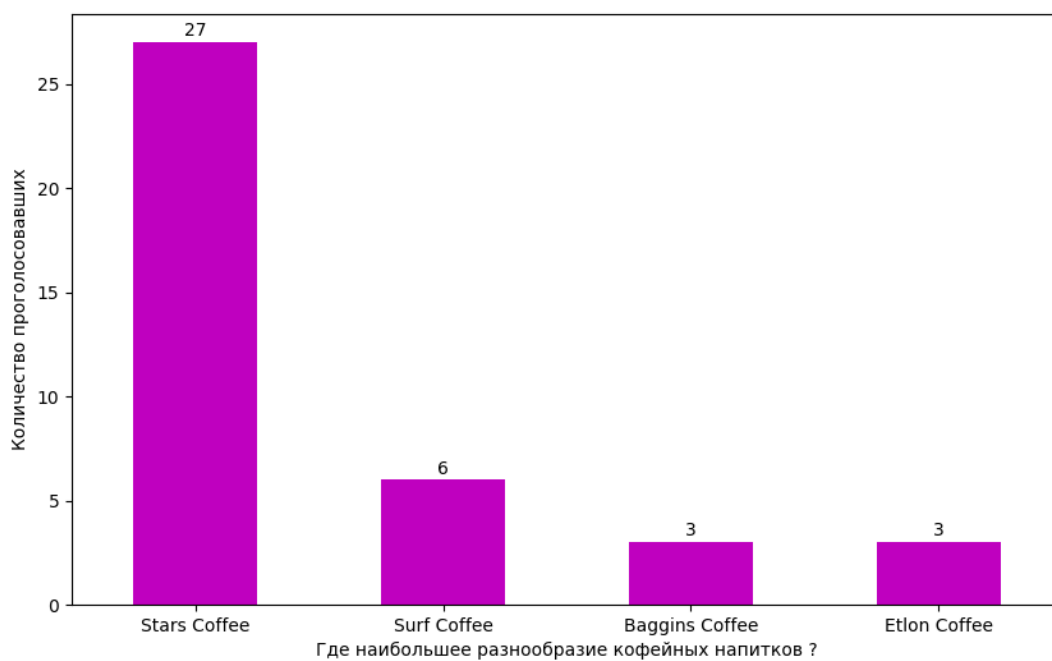
Далее отображаем результаты опроса в виде графиков:

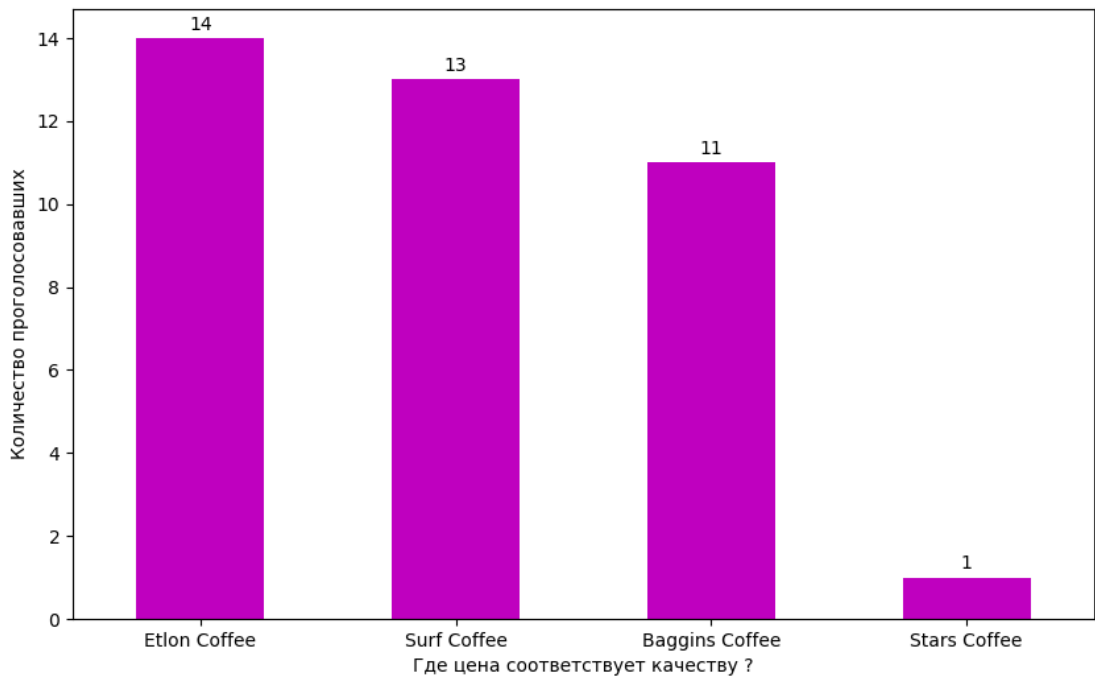
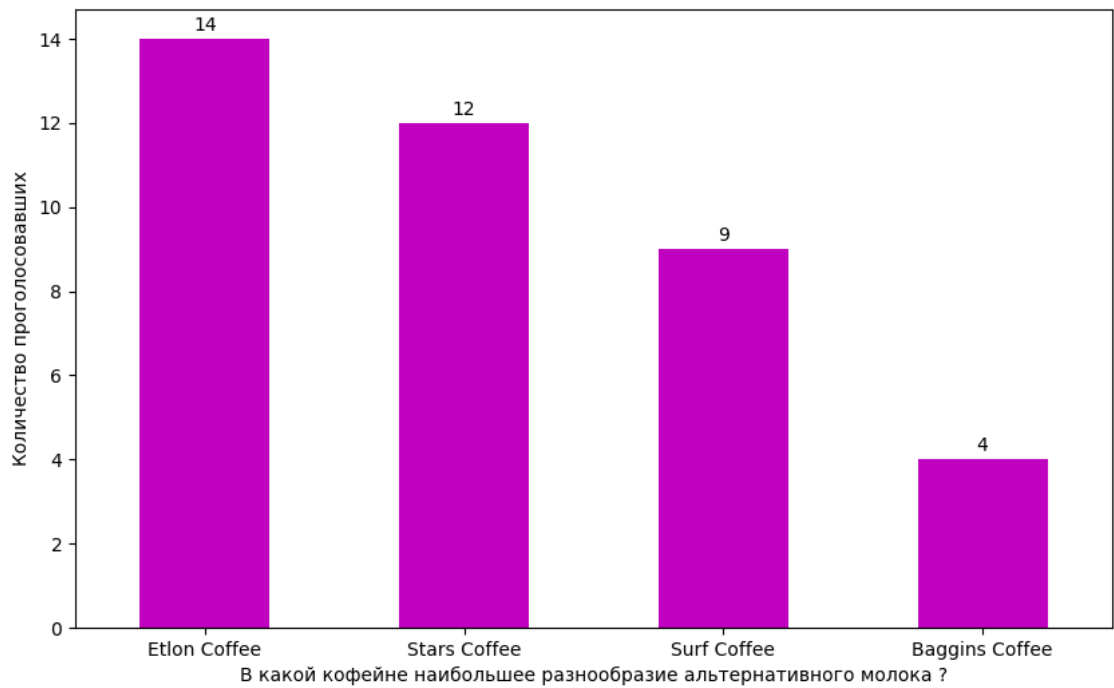
```
# Функция для построения столбчатого графика
def plot_bar_chart(column_name):
    fig, ax = plt.subplots(figsize=(10, 6))
    value_counts = df[column_name].value_counts()
    df[column_name].value_counts().plot(kind='bar', color='m', ax=ax, rot =0)
    plt.xlabel(column_name)
    plt.ylabel('Количество проголосовавших')
    for i, v in enumerate(value_counts):
        ax.text(i, v + 0.1, str(v), ha='center', va='bottom', fontsize=10)

    plt.show()

# Построение графика для каждой оценки
for column in df.columns[2:]:
    plot_bar_chart(column)
```

Графики:





Далее проведем исследование и вычислим коэффициент V Крамера на основе таблицы сопряженности. Проведем хи-квадрат тест и рассчитаем коэффициент V Крамера:

```
# Преобразование категориальных переменных в числовые для анализа
df_numeric = df.copy()
df_numeric.replace({"Baggins Coffee": 1, "Etlon Coffee": 2, "Stars Coffee": 3, "Surf Coffee": 4}, inplace=True)

# Создание всех возможных комбинаций переменных
variable_combinations = list(combinations(df_numeric.columns[2:], 2))

# Проведение анализа для каждой пары переменных
for variable1, variable2 in variable_combinations:
    # Создание таблицы сопряжённости
    contingency_table = pd.crosstab(df_numeric[variable1], df_numeric[variable2])

    # Проведение хи-квадрат теста
    chi2, _, _, _ = chi2_contingency(contingency_table)

    # Расчет коэффициента V Крамера
    n = contingency_table.sum().sum()
    v_cramer = (chi2 / (n * (min(contingency_table.shape) - 1))) ** 0.5

    # Вывод результатов для каждой пары переменных
    print(f"Коэффициент V Крамера для '{variable1}' и '{variable2}': {v_cramer:.3f}\n")
```

В результате получили результаты:

```
Коэффициент V Крамера для 'Где наибольшее разнообразие кофейных напитков ?' и 'Где кофе наиболее крепкий ?': 0.290
Коэффициент V Крамера для 'Где наибольшее разнообразие кофейных напитков ?' и 'В какой кофейне наибольшее разнообразие альтернативного молока ?': 0.335
Коэффициент V Крамера для 'Где наибольшее разнообразие кофейных напитков ?' и 'Где цена соответствует качеству ?': 0.324
Коэффициент V Крамера для 'Где кофе наиболее крепкий ?' и 'В какой кофейне наибольшее разнообразие альтернативного молока ?': 0.394
Коэффициент V Крамера для 'Где кофе наиболее крепкий ?' и 'Где цена соответствует качеству ?': 0.308
Коэффициент V Крамера для 'В какой кофейне наибольшее разнообразие альтернативного молока ?' и 'Где цена соответствует качеству ?': 0.371
```

Коэффициент V Крамера используется для измерения силы ассоциации между двумя категориальными переменными. Здесь мы в качестве переменных использовали вопросы, превратим результаты в таблицу для наглядности.

1 вопрос	2 вопрос	Коэффициент V Крамера
Разнообразие кофейных напитков	Крепкий кофе	0.290
Разнообразие кофейных напитков	Разнообразие альтернативного молока	0.335
Разнообразие кофейных напитков	Цена == качество	0.324
Крепкий кофе	Разнообразие альтернативного молока	0.394
Крепкий кофе	Цена == качество	0.308
Разнообразие альтернативного молока	Цена == качество	0.371

Зухир Амира Саидовна, 3 курс ИВТ, группа 1.2

Значения коэффициента V Крамера находятся в диапазоне от 0 до 1, где:

- 0: отсутствие ассоциации
- 1: полная ассоциация

Так как наши все значения **ближе к нулю, чем к единице**, значит переменные (вопросы в нашем случае) **независимы друг от друга**.

Код с файлом json и всеми графиками доступен по [ссылке](#).