

# Отчет по анализу меню ресторанов быстрого питания



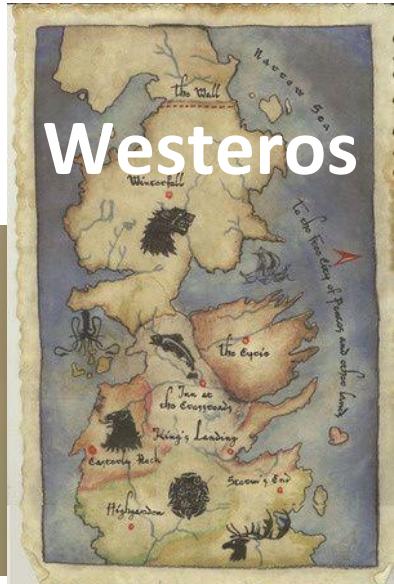
VS



◆ Aleksandr Berezkin

◆ Merr

◆ Fatima Jandossova



Отчет подготовлен командой аналитиков  
«Westeros»

для детского развивающего центра  
«Мега Мозг»

Westeros - 2019

# Задачи анализа

**Провести анализ меню McDonalds (далее - McD) и Starbucks (далее - SB) для заключения о том, есть ли возможность подобрать в этих ресторанах меню, пригодное для питания подростков.**

Требования: меню должно быть сбалансировано по содержанию основных компонентов пищи, содержать достаточное, но не избыточное количество калорий и – по возможности – позволять комбинировать разные блюда в зависимости от вкуса конкретного ребенка.

- **Техническая задача**

Извлечь из наборов данных нужную информацию и представить её в виде отчёта о соответствии меню ресторанов McDonalds и Starbucks требованиям здорового питания



# План исследования

1. Изучить предоставленные данные, понять полноту и достаточность, выбрать наилучшие данные для соответствия задаче
2. Продумать логику сбора и анализа данных, оценить возможность автоматизации сбора, обработки и анализа данных
3. Собрать и изучить основные и дополнительные показатели
4. Разработать иллюстрации, отражающие особенности меню двух ресторанов и соответствие меню правилам здорового питания.
5. Разработать общее заключение и рекомендации по организации питания подростков в ресторанах быстрого питания.
6. Подготовить отчёт о проведённом анализе. Отчёт должен включать текстовую, графическую и табличную информацию
7. Загрузить результаты работы (текстовый отчёт и ноутбук с кодом и графиками) на GitHub (<https://github.com>) и предоставьте ссылку

# Обоснование избранного подхода



Критерии  
Наличие, присутствие в  
меню

# Основная идея поиска решения

**Комбинаторный подход:** создание «осмысленных» комбинаций из меню каждого из ресторанов с последующим выбором комбинаций, соответствующих определенным критериям.

- Наборы должны содержать 2 или 3 блюда, без повторов
- Наборы не должны содержать только еду или только напитки
- Алгоритм должен последовательно возвращать все возможные комбинации, а не случайные

# Преимущества

- Подход позволяет не просто ответить на вопрос «Какой ресторан лучше», но и предложить **максимально разнообразное меню, удовлетворяющее критериям сбалансированного питания**, вне зависимости от выбранного ресторана.
- Критерии (как статистические, так и по отдельным показателям) можно гибко задавать, менять, быстро получать результаты, сохранять и сравнивать их.
- Подход может рассматриваться как задел, proof of concept для дальнейшего развития решения.

# Первоначальная аналитика



Обработка данных и исследовательский  
анализ

Полнота и достаточность данных  
Логика сбора и анализа

Оценка возможности автоматизации сбора,  
обработка и хранение

# Исходные данные: неполнота

Был проведен анализ предоставленных заказчиком данных о блюдах по 6 указанным показателям, а также данных с сайтов компаний McD и SB

1. Наборы данных в csv: питательные вещества и энергетическая ценность, которые содержатся в различных блюдах и напитках обеих компаний, неполные;
2. Сайты: цены и наборы блюд (завтраки/обеды)
3. Время года: начало лета (июня)
4. **Регулярность** выезда детей: “ученики школы **регулярно** выезжают на экскурсии и **спортивные** мероприятия, которые часто делятся 3-4 часа и **более**”
5. Не все данные присутствуют в меню компаний. **Найдены только 6 показателей для сравнения** по питательным веществам.
6. Ни в одном ни в другом наборе данных нет данных по аллергенам (есть отдельно на сайтах).
7. Наборы данных меню от клиента частично устарели (по SB)

# Какой информации не хватает?

В представленных данных нет информации о **нормативах по питанию подростков**. Чтобы определить, **пригодно ли меню для питания подростков**, надо ответить на вопросы:

1. Какое меню вообще пригодно для подростков? Есть отличия от взрослых? Какие блюда выбрать?
2. Какой возраст считать **подростковым**?
3. Нужно ли разделять меню отдельно для **мальчиков и девочек**?
4. Есть ли специальные нормативы для **спортивного питания**? Насколько нагрузки будут соответствовать спортивным и будут велики?
5. Нужно ли менять наборы блюд **ежедневно**? И как часто надо менять блюда?

# Допущения для формирования базы

После ознакомления с нормативами (см. список источников) сделаны допущения:

1. Клиент находится (или работает) в России. Можно использовать соответствующие [нормативные документы](#) по детскому питанию (Табл. 5.4)
2. Подростковый возраст - [от 14 до 18 лет](#) (п. 3.1)
3. Использованы общие нормативы для мальчиков и девочек: взяты усредненные цифры, где они различаются.
4. Потребность в натрии увеличивается при спортивной активности: норма в 0,78 г была обновлена до 1,0 г для натрия. В частности, это позволило сохранить больше элементов в наборе данных.
5. По всем показателям можно было бы подобрать “усиленные рационы” для спортивных дней: +20% к норме.
6. Дети посещают лагерь каждый день (5-дневка). Максимально, в течение 5 дней еда не должна повторяться за два последних дня (как минимум **три разных набора блюд для завтраков и аналогично для обедов**).

# Выбор показателей



Критерии  
Наличие, присутствие в  
меню

# Показатели для оценки меню ресторанов

- Энергетическая ценность = **Энергия**
- Общее содержание жиров = **Жиры**
- Общее содержание углеводов = **Углеводы**
- Общее содержание пищевых волокон = **Клетчатка**
- Общее содержание белков = **Белки**
- Общее содержание натрия = **Натрий**

Только эти 6 показателей были найдены в обоих наборах данных по ресторанам, а также присутствовали в нормативах. Мы их взяли за основу для оценки сбалансированности состава продуктов конкретного варианта завтрака / обеда.

**Показатели присутствуют во всех трех наборах данных (нормы, McD и SB)**

<b>Наименование</b>	<b>Англ. наименование</b>	<b>Ед. измерения</b>	<b>Значение дневной нормы</b>
Энергия:	Energy	(kcal)	2900
Жиры:	Fat	(g)	97
Белки:	Protein	(g)	87
Углеводы:	Carbohydrates	(g)	421
Клетчатка:	Fiber	(g)	20
Натрий:	Sodium	(g)	1.3

# Объединенная база 500 блюд - основные статистики

	Energy(kcal)	Fat(g)	Carbohydrates(g)	Fiber(g)	Protein(g)	Sodium(g)
count	500.00000	500.000000	500.000000	500.000000	500.000000	500.000000
mean	302.72000	11.961000	38.740000	1.841000	10.776000	0.389510
std	212.84974	12.144338	25.376215	2.325087	10.325457	0.495669
min	0.00000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	140.00000	1.500000	21.750000	0.000000	2.000000	0.058750
50%	270.00000	9.000000	36.000000	1.000000	9.000000	0.162500
75%	430.00000	19.000000	52.000000	3.000000	16.000000	0.590000
max	1880.00000	118.000000	141.000000	24.000000	87.000000	3.600000

# Примечание по допущению относительно гендера



Поскольку нормы потребления по энергии, жирам, углеводам и белкам для мальчиков больше (примерно на 16%), усредненные значения могут вызывать недостаточные уровни потребления для мальчиков (-8%) и избыточные для девочек (+8%).

Это можно учесть и реализовать дополнительные сценарии фильтрации показателей с учетом различий норм для девочек и мальчиков.

# Нормирование



Согласно [рекомендациям](#) по питанию в детских учреждениях на завтрак приходится 25%, а на обед 35% дневной нормы калорийности и питательных веществ.

В задании речь идет **только о завтраках и обедах**. Мы взяли 30% от дневной нормы, то есть для завтрака / обеда = норма \* 0,3

Неповторяемость меню в течение дня / двух дней подряд  
(должно быть более 1 набора в 1 ресторане).

# Подготовка данных



Единая система измерений  
Исключение неполных  
данных

# Подготовка базы данных

1. Данные из McD.csv and SB.csv были объединены в Excel (добавление признаков в доп. колонки)
2. Добавленные признаки: (1) «McD / SB»: кодировка ресторана, (2) «Категория»: еда / напитки, (3) «Вид»: определенная группа товаров, например: «Выпечка», «Горячий завтрак», «Сладости», «Горячий напиток». У SB таких групп позиций примерно в 2 раза больше, чем у McD, но у последних в каждой группе больше блюд (напитков).
3. Набор данных был также обогащен вручную: > 60 единиц в категории «Закуски и сладости» были добавлены в меню SB: набор данных выровнялся по представленности каждого ресторана
4. Все данные по пищевой ценности переведены для в граммы
5. Удалены некоторые символы в названиях блюд (©, ™ и т.д.)
6. Удалены позиции, где все значения нулевые, позиции, которые превышали 60% дневной нормы хотя бы по одному показателю.

**без спортивной  
нормы?**

# Сравнительный анализ меню ресторанов по питательным веществам

Визуализация и исключение  
некоторых блюд



# Посчитаны основные статистики



mean  
std  
min  
25%  
50%  
75%  
max



По каждому из показателей (**Energy, Fat, Carbohydrates, Fiber, Protein, Sodium**) в разбивке по ресторанам, получены соответствующие визуализации.

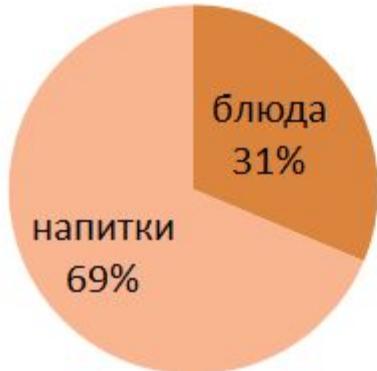
Результаты анализа обобщены в виде комментариев в Jupyter Notebook и описаны в разделе результатов.

## Соотношение напитков и блюд

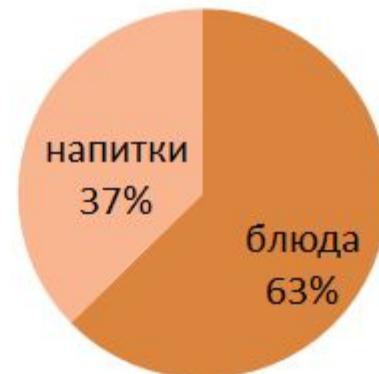
У SB гораздо больше продуктов питания (всего 136 из 217), в то время как у McD гораздо больше напитков (всего 138 из 201)



McD

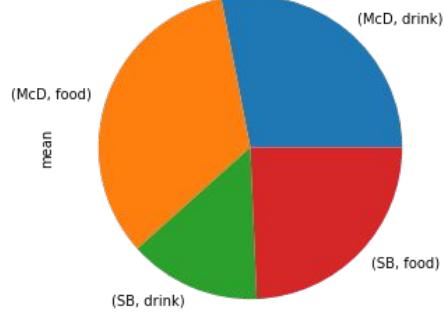


SB

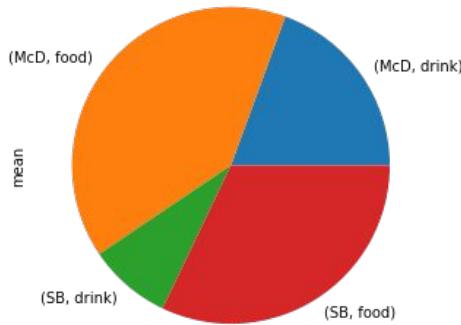


# Общий анализ по питательным веществам McD vs SB: напитки и еда

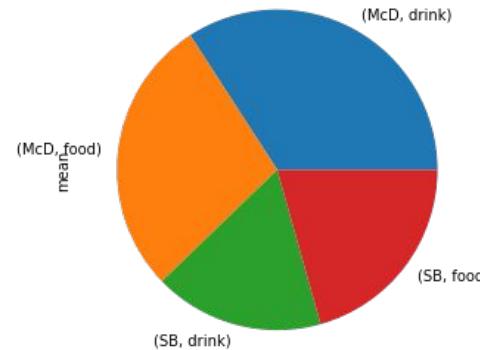
## Энергия



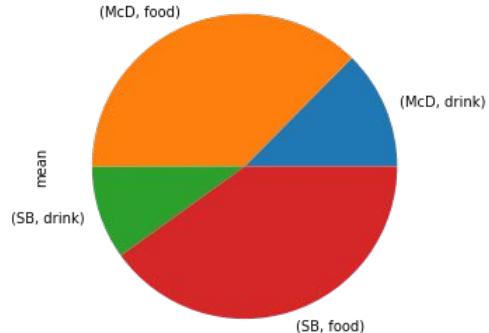
## Жиры



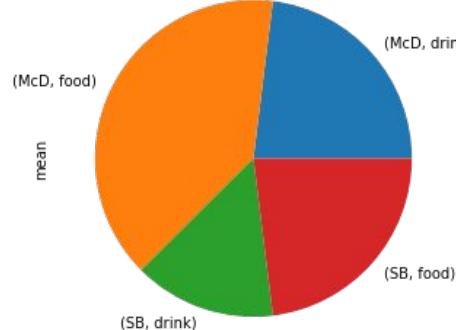
## Углеводы



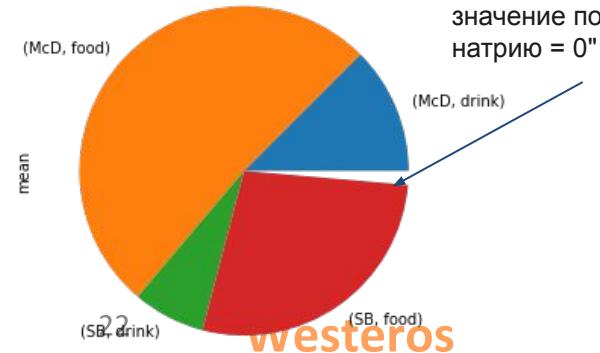
## Клетчатка



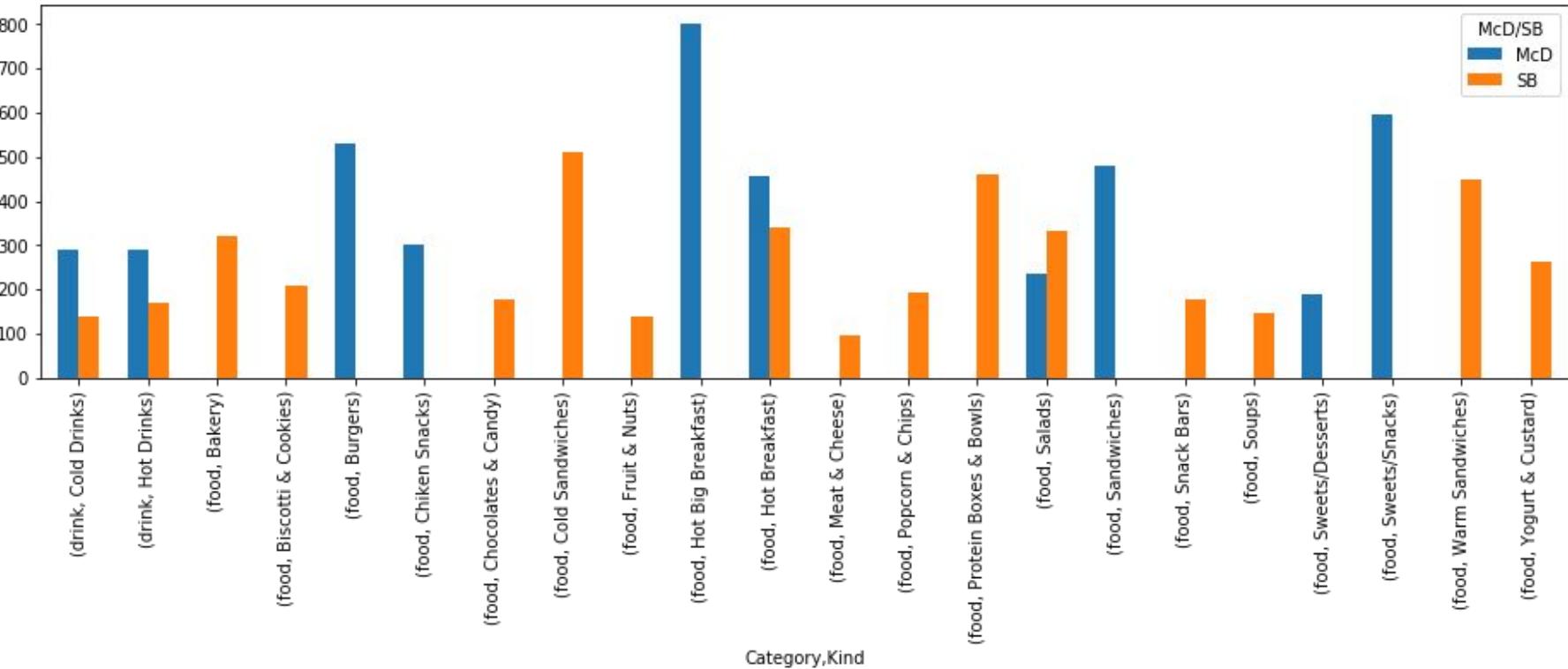
## Белки



## Натрий

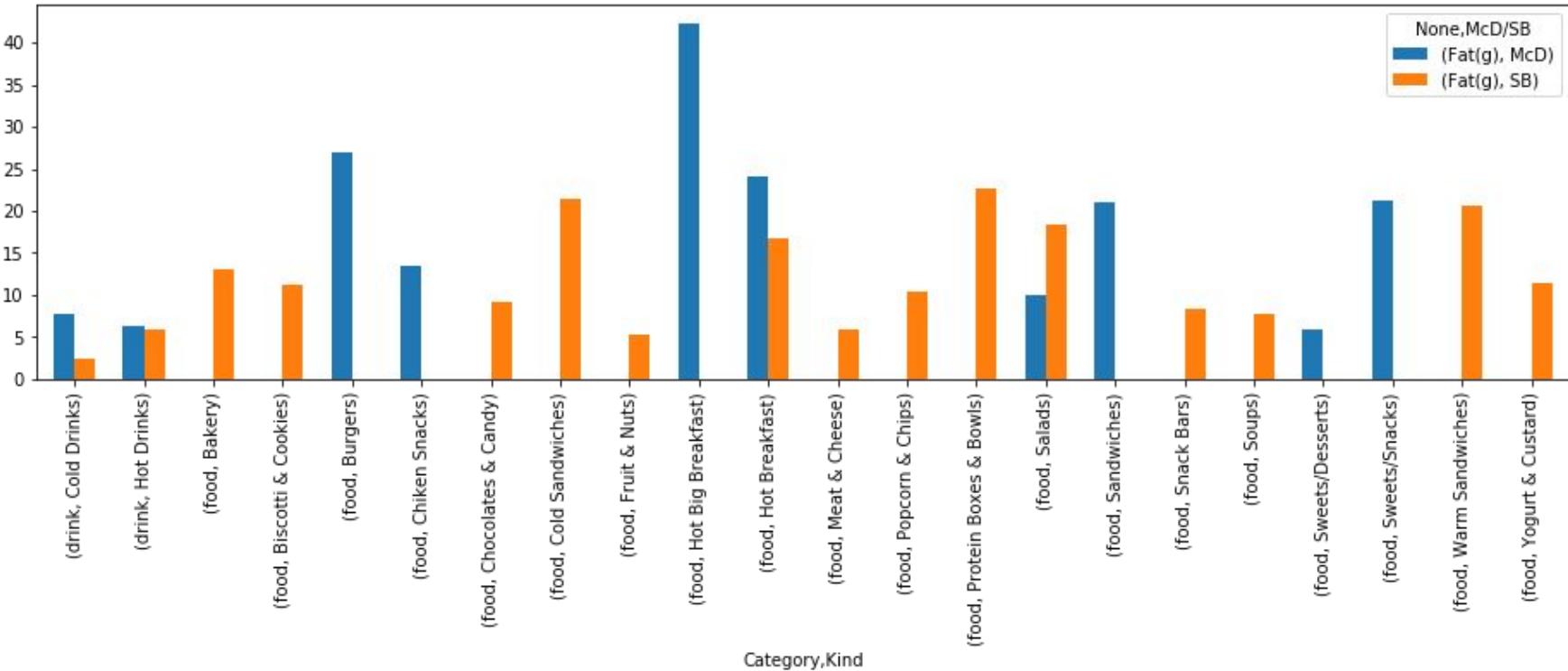


# Разнообразие меню и его калорийность



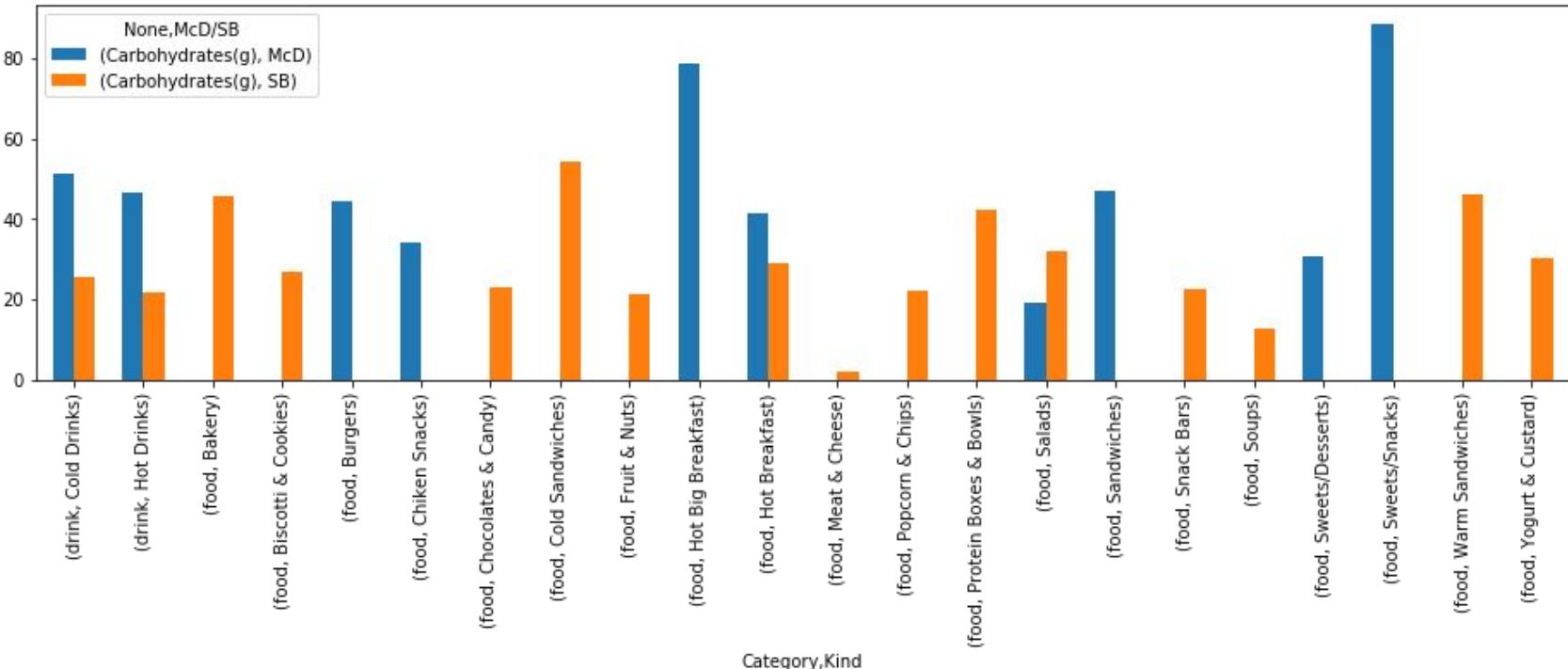
Мы видим, что в SB менее калорийные блюда в целом, и их разнообразие выше по категориям блюд, чем в McD

# Содержание жиров



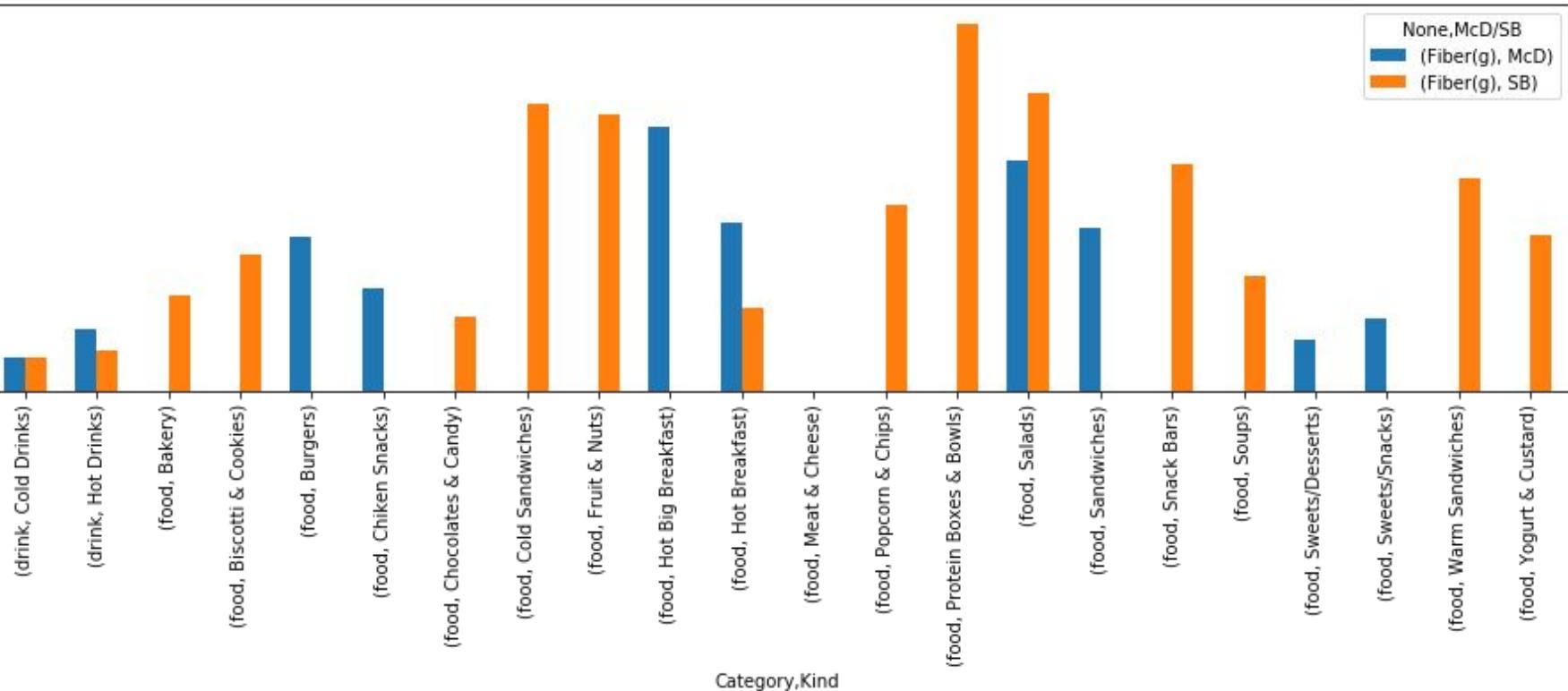
В McD в холодных напитках, горячем завтраке больше жиров, а в салатах - меньше жиров, чем в меню SB

# Содержание углеводов



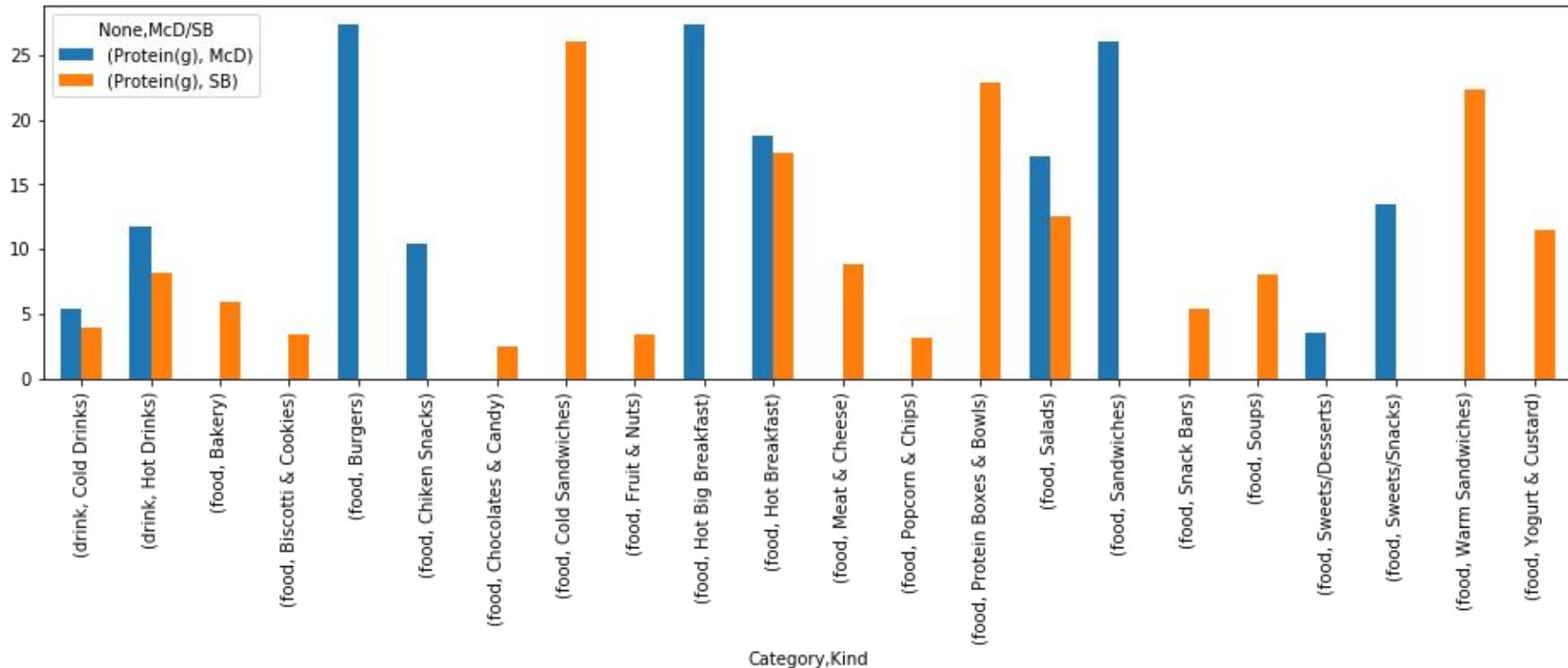
В SB количество углеводосодержащих блюд и напитков в целом меньше, чем в McD

# Содержание клетчатки



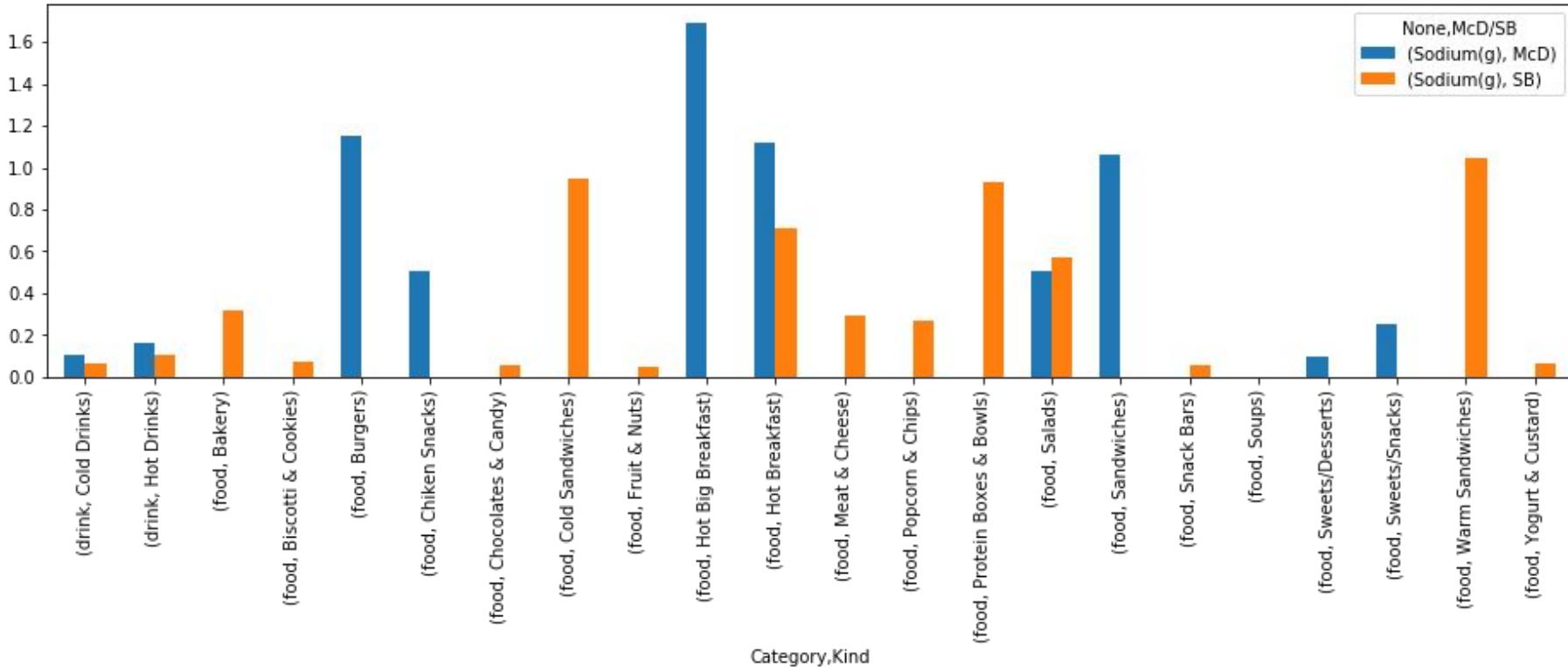
Мы видим, что в SB содержание клетчатки заметно выше, чем в McD!

# Содержание белков



McD по всем сравнимым категориям содержит больше белков, чем в SD!

# Содержание натрия



McD чуть больше содержит натрия, а в горячих завтраках особенно, чем в SD.  
Иключение - салаты. Если сравнивать бургеры и сэндвичи, то примерно равное количество.

# Комбо-меню

Сочетания наборов блюд



# Алгоритм анализа комбо-меню

1. Нормы: чтение в Pandas в виде словаря meal\_norm.
2. Загрузка словаря со взвешенной нормой (в расчете на один прием пищи), нормирование значений на основе данных словаря: показатели блюд выражены в долях от нормы в расчете на один приём пищи.
3. Выделение меню каждого ресторана из общего датасета. Для удобства работы по каждому меню применялся отдельный ноутбук с одним и тем же алгоритмом.
4. Разработка форматов завтраков / обедов: «1-ое блюдо, 2-ое блюда, 3-е (напиток)» и др.
5. Составление вручную набора групп 1-х, 2-х блюд и напитков для завтраков и обедов. Категории выбирались в соответствии с названиями («Hot Breakfast» в меню обоих ресторанов включался в состав завтраков; Lunch/Sandwiches в меню SB – в состав обедов и т.д.), а также представлениями о классическом завтраке / обеде, например, супы – на обед, салаты – как первые блюда на завтрак или обед и т.п.

# Два типа набора блюд

1 тип  
набор  
а



1 блюдо  
2 блюдо  
и  
напиток



2 тип  
набор  
а



1 блюдо  
и  
напиток

# Алгоритм составления вариантов

1. Реализация алгоритма, составляющего бесповторные поэлементные комбинации из элементов наборов 1-х, 2-х блюд и напитков., а также завтраков / обедов других форматов. Результат: датасеты комбинаций блюд для завтраков и обедов в 5 форматах для каждого из ресторанов – всего 10 датасетов.
2. Расчет суммарных значений для каждой комбинации блюд по каждому из 6 нормируемых параметров. Например, одна из комбинаций состоит из супа, горячего сэндвича и кофе: для всех этих продуктов суммируются содержания калорийности, белков, жиров... и т.д для всех полученных комбинаций.
3. Расчет средних значений и стандартных отклонений по 6 суммарным показателям каждой комбинации блюд.
4. Сборка всех полученных комбинаций с посчитанными метриками в единый датасет результатов.

# Результаты



Результаты анализа  
сгенерированных вариантов  
для завтраков / обедов. Выбор  
лучших вариантов

## Среднее и увеличение нормы на 20%

- ❖ В результате применения комбинаторного алгоритма было сгенерировано **более 360 тыс вариантов** для завтраков / обедов, состоящих как на основе двух блюд, так и на основе 3 блюд.
- ❖ Далее мы применили несколько сценариев выбора лучших блюд из полученного набора: отфильтровали варианты с определенными средними значениями (MEAN) и стандартными отклонениями (STDEV).
- ❖ Среднее значение = 1 соответствует норме по данному показателю в расчете на один прием пищи, среднее – 1,2 соответствует увеличению нормы на 20% и т.д.

# Варианты рамок отбора

## “Жесткая норма”

MEAN = 1+/-0.1, STDEV <= 0.1

```
mega_data_best =  
    mega_data[(mega_data['  
        MEAN']) > 0.9) &  
        (mega_data['MEAN'] <  
        1.1)]
```

298 –  **571**  
вариант  
ов

## “Мягкая норма”

MEAN = 1+/-0.25, STDEV

```
mega_data_best =  
    mega_data[(mega_data['  
        MEAN']) > 0.75) &  
        (mega_data['MEAN'] <  
        1.25)]
```

2602 –  **5538**  
вариант  
ов

## “Спортивный день”

MEAN = 1.2+/-0.1, STDEV <= 0.1,

```
Energy > 1  
mega_data_best =  
    mega_data[(mega_data['  
        MEAN']) > 0.9) &  
        (mega_data['MEAN'] <  
        1.1)]
```

996 –  **1002**  
варианта  
*Westeros*



# Заключение

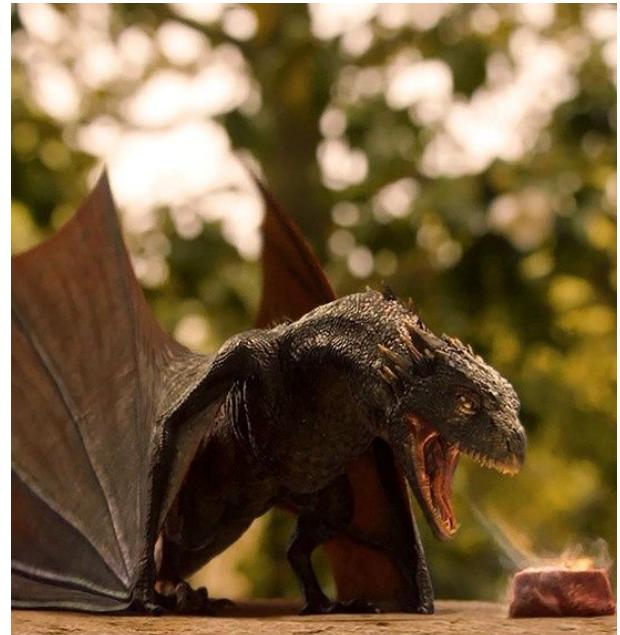
Какой ресторан выбрать?

## Общее заключение

- Если клиент ставит приоритетом **сбалансированное питание**, которое соответствует нормам, то разумнее будет выбрать меню McDonalds.
- Если же значения нормы не столь важны (но все же учитываются) - меню Starbucks будет гораздо предпочтительнее (гораздо больше блюд этого ресторана проходят более "мягкий" отбор по нормам).
- Когда вам нужна высококалорийная и богатая питательными веществами пища для обеспечения питания в спортивные дни, разумнее выбрать меню McD

# Примечание

- ❖ Полученные "лучшие комбинации" на основе меню обоих ресторанов – не более чем результат работы алгоритма и фильтрация данных по формальным критериям.
- ❖ Мы надеемся, что этих списков достаточно, чтобы обеспечить разнообразное меню на довольно длительный срок (независимо какой ресторан клиент выберет в итоге).
- ❖ Большой набор данных позволяет выбрать различные варианты, полностью соответствующие запросам клиента (например, с учетом аллергенов в продуктах меню).



# Ответы на ключевые вопросы задачи

- ❖ Можно ли составить меню, которое было бы сбалансировано по содержанию основных компонентов пищи, содержать достаточное, но не избыточное количество калорий?

Да, это возможно

- ❖ Можно ли составить меню, которое позволило бы комбинировать разные блюда в зависимости от вкуса конкретного ребенка?

Да, это возможно с большой долей вероятности, поскольку комбинаций блюд множество.

Однако, количество комбинаций для завтраков в "жестком" меню ограничено для McD,

"Жесткая норма" - для SB. Ниже представлено количество комбинаций по "Мягкой норме" "Спортивный день" вариантам:

McD_breakfast	1	McD_breakfast	11	McD_dinner	595
McD_dinner	217	McD_dinner	1508	McD_light_dinner	371
McD_light_breakfast	6	McD_light_breakfast	39	SB_breakfast	4
McD_light_dinner	62	McD_light_dinner	973	SB_dinner	32
McD_snack	12	McD_snack	71		
SB_breakfast	27	SB_breakfast	515		
SB_dinner	196	SB_dinner	2410		
		SB_light_breakfast	2		
		SB_light_dinner	9		

## Доп. рекомендации для клиента

- **Приоритет горячих блюд.** Например, Warm sandwiches вместо Cold sandwiches в меню SB, Hot drinks вместо Cold drinks в обоих меню.
- **Учет уровня сахара в продуктах.** Особенно это касается очень калорийных и богатых сахаром мороженых из меню McD.
- **Контроль уровня соли.** Ранее мы увеличили норму по содержанию натрия с 0.78 до 1.0 г из расчета на один прием пищи, поскольку это позволяло оставить больше блюд в исходном наборе. Мы считаем, что это приемлемо, т.к. в при физических нагрузках организм теряет натрий. Однако необходимо уделить особое внимание уровню соли в продуктах, кроме того, мы рекомендуем обеспечить доступ детей к питьевой воде не только при приеме пищи, но и в процессе спортивных мероприятий.

# Предложения по дальнейшей работе аналитиков

1. Проверка других сценариев с различными фильтрами. Фильтры могут быть детализированы "по требованию" клиента.
2. Включите данных о размерах порций и потенциальных аллергенах в продуктах в качестве новой функции и дополнительных фильтров, например, исключение супов.
3. Поиск индексов питания или расчет интегрального индекса питания. Например, оптимальное белков, жиров и углеводов = 1:1:4 или в % от калорийности: 10-15%, 30-32% и 55-60% соответственно. Применение подобных индексов к полученному массиву комбинаций на основе исходных меню.
4. Улучшение кода. Пересмотр кода (более "питонический") и полная автоматизация алгоритма анализа в одном ноутбуке. Обертывание всех процедур в соответствующие функции или классы.
5. Методы ML вместо ручных "сценариев": мы считаем, что методы оптимизации подойдут для данной задачи.
6. Генератор меню с различными настройками фильтров с GUI. Можно загружать пользовательские датафреймы, получать список комбинаций и устанавливать значения критериев для выбора лучших. Если список одобрен пользователем, из отобранных в результате вариантов программа может выбрать N случайных для составления уникальных меню на много дней. Это обеспечит разнообразие блюд, которого хватит надолго.

# Источники

и др. информация



# Список источников. Начальные данные

**Официальные данные** о блюдах и их компонентах, предоставленные компаниями McDonald's и Starbucks

- Nutrition Facts for McDonald's

Menu: <https://www.kaggle.com/mcdonalds/nutrition-facts>

- Nutrition facts for Starbucks

Menu: <https://www.kaggle.com/starbucks/starbucks-menu#starbucks-menunutrition-food.csv>

**Дополнительная информация о меню**, составе продуктов и ценах ресторанов:

- McDonald's: <https://www.mcdonalds.com/us/en-us/full-menu.html>

или <https://mcdonalds.ru/products>

- 

Starbucks: <https://www.starbucks.com/menu> или <http://www.starbucks.ru/menu>

# Список источников. Сбалансированное питание:

- Методические рекомендации МР 2.3.1.2432-08 Нормы физиологических потребностей в энергии и пищевых веществах для различных групп населения Российской Федерации  
<https://base.garant.ru/2168105/>, <http://www.normacs.ru/Doclist/doc/11HIP.html>
- Сбалансированное питание на сайте MD <https://mcdonalds.ru/page/balanced/>
- Меню на сайте SB [https://www.starbucks.com/menu#view\\_control=nutrition](https://www.starbucks.com/menu#view_control=nutrition)
- Здоровая диета для подростка-мальчика  
<https://www.everydayme.ru/dom-i-sad/gotovim-domu/pravilnoe-pitanie-dlya-malchikov-ot-desyati-let-i-podrostkov>
- МР 2.4.5.0131-18. 2.4.5. Гигиена детей и подростков. Детское питание. Практические аспекты организации рационального питания детей и подростков, организация мониторинга питания. Методические рекомендации" (утв. Главным государственным санитарным врачом РФ 10.08.2018)  
<https://legalacts.ru/doc/mr-2450131-18-245-gigiena-detei-i-podrostkov-detskoe-pitanie/>
- Нормы по спортивному питанию подростков  
<http://demo.garant.ru/#/document/70170148/paragraph/874/doclist/0/seflink/0/highlight/%D0%A1%D0%BF%D0%BE%D1%80%D1%82%D0%B8%D0%B2%D0%BD%D0%BE%D0%B5%20%D0%BF%D0%B8%D1%82%D0%B0%D0%BD%D0%B8%D0%B5%20%D0%BF%D0%BE%D0%B4%D1%80%D0%BE%D1%81%D1%82%D0%BA%D0%BE%D0%B2:0>
- Письмо Министерства образования и науки РФ от 12 апреля 2012 г. N 06-731 "О формировании культуры здорового питания обучающихся, воспитанников"  
<https://legalacts.ru/doc/pismo-minobrnauki-rossii-ot-12042012-n-06-731/>

# Westeros



## Задачу решали

**Merr:** Работа с дадасетами в Pandas - написание кода + визуализация + документирование кода и всех шагов + описание аналитики, заключение.

**Aleksandr Berezkin:** алгоритмы, обсуждения

**Fatima Jandossova:** поиск норм, оформление отчета