

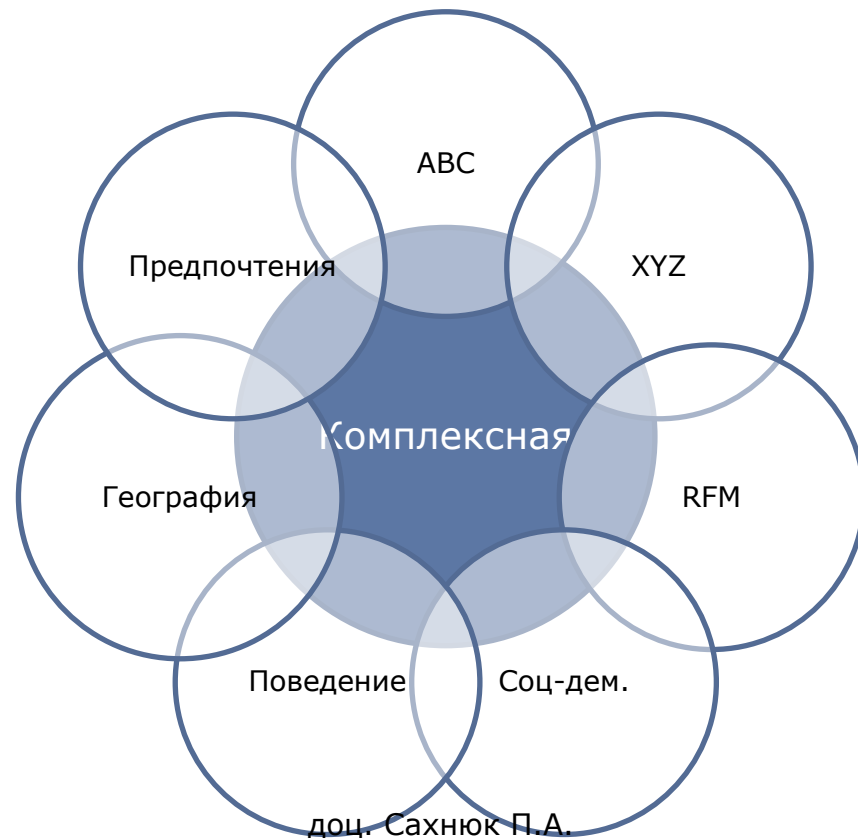
# Маркетинговая аналитики на SQL

доц. Сахнюк П.А.

# Вопросы к аналитике

- Кого привлекать?
- Как удержать ценных клиентов?
- Как увеличить прибыльность?
- Как сформировать привлекательные адресные предложения?
- Как увеличить отклик на предложения?
- Как минимизировать отток?
- Как минимизировать негатив?
- Как диагностировать проблемы?

# Варианты анализа



# Простые методы

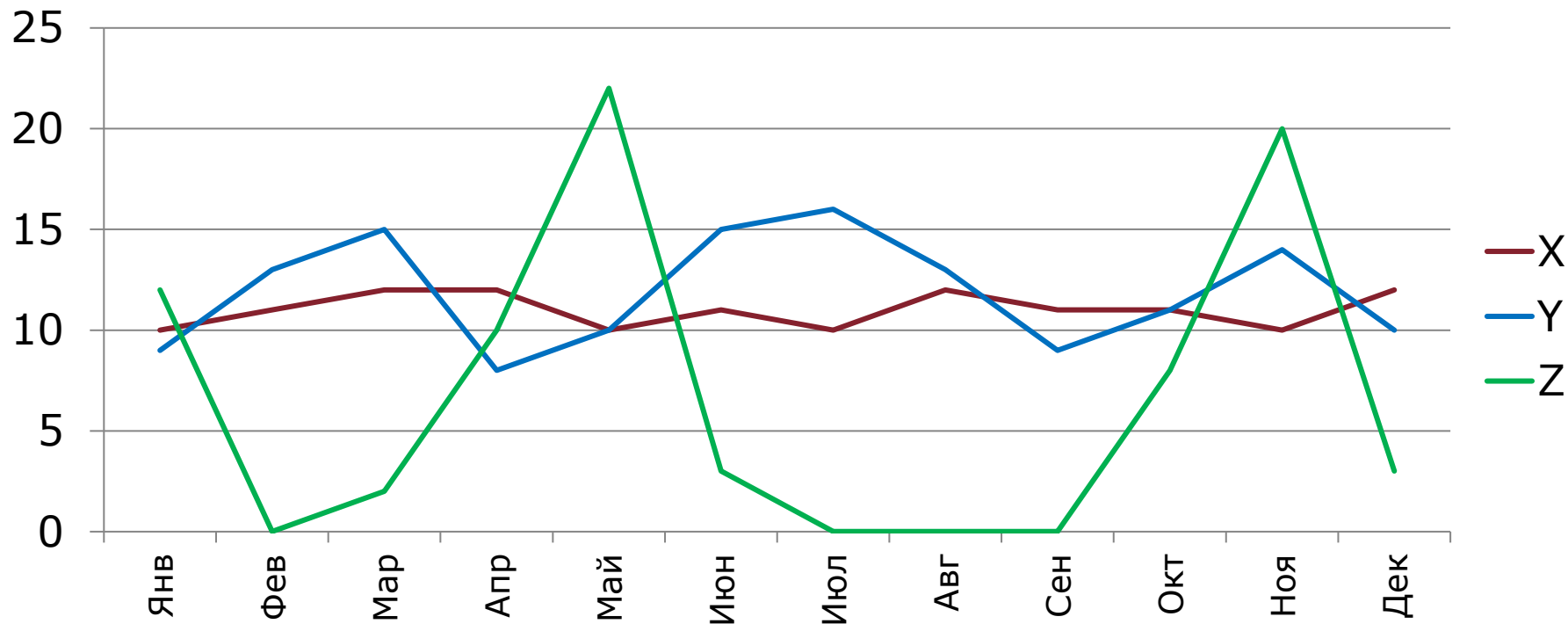
## ABC-анализ: алгоритм

1. Выбрать показатель: выручка, маржа...
2. Агрегировать данные по каждому клиенту за последний период
3. Отсортировать по убыванию показателя, рассчитать накопительную сумму:
  - Первые 20% - самые ценные
  - Следующие 30% - промежуточные
  - Последние 50% - наименее ценные

# XYZ-анализ: алгоритм

1. Выбрать квант: день, неделя, месяц
2. Агрегировать данные по продажам каждому клиенту и получить временные ряды
3. Рассчитать коэффициент вариации для каждого клиента:
  - 0-10% - стабильное потребление
  - 10-25% - регулярное потребление
  - >25% - хаотическое потребление

# XYZ-анализ: пример



доц. Сахнюк П.А.

# RFM-анализ: идея

Оценка лояльности клиента на основе показателей:

- **R**ecency (новизна) – давность последней покупки.
- **F**requency (частота) – частота покупок.
- **M**onetary (деньги) – потраченные суммы

# RFM-анализ: алгоритм

Каждый из показателей делится на 5 квантилей, например:

R	Давность покупки
5	<1 мес.
4	2-3 мес.
3	4-8 мес.
2	9-18 мес.
1	>18 мес.

F	Частота покупок
5	>1 в 1 мес.
4	1 в 2-3 мес.
3	1 в 4-8 мес.
2	1 в 9-18 мес.
1	1 в >18 мес.

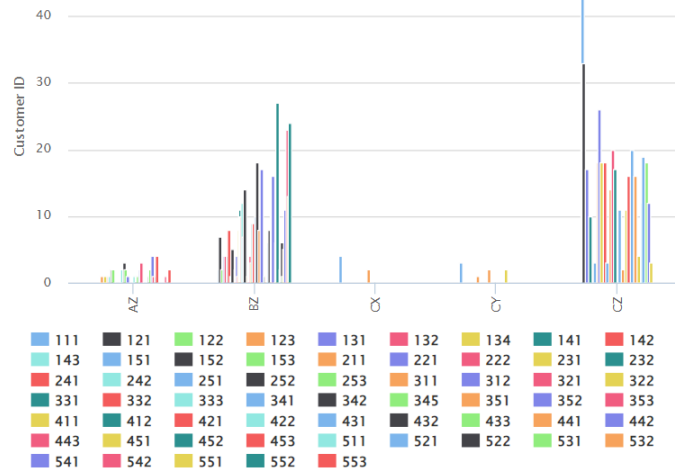
M	Потраченные суммы
5	>100 000 руб.
4	50-100 000 руб.
3	10-50 000 руб.
2	3-10 000 руб.
1	<3 000 руб.

Считается RFM-код каждого клиента.



# ABC-XYZ, RFM-анализ: визуализация

RFM	ABC-XYZ	AZ	BZ	CX	CY	CZ	Итого:
443		1					1
451			6			4	10
452		1	27				28
453		4					4
511						12	12
521			2			19	21
522			6				6
531			1			18	19
532			5				5
541			11			12	23
542		1	23				24
551			13			3	16
552		2	24				26
553		2					2
Итого:		42	353	6	8	384	793



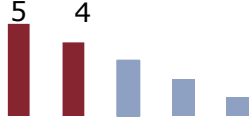
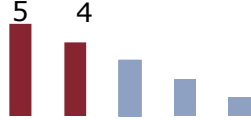
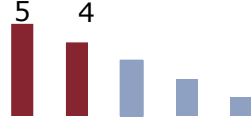

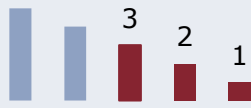

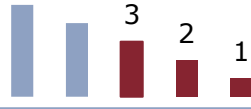

#	ab Customer...	ab Order_ID	31 Order_D...	ab Customer Name	ab Segment	90 Sales	12 Quantity	ab ABC-XYZ	31 Order_Date Максимум	12 Order_ID Количество	31 Order_Date Максимум	12 Order_ID Ко
1	SE-20110	US-2014-121734	11.04.2017	Sanjit Engle	Consumer	9,584	1	AZ	21.12.2020	19	21.12.2020	
2	SE-20110	CA-2014-158225	22.09.2017	Sanjit Engle	Consumer	169,45	5	AZ	21.12.2020	19	21.12.2020	
3	SE-20110	CA-2014-158225	22.09.2017	Sanjit Engle	Consumer	40,68	2	AZ	21.12.2020	19	21.12.2020	
4	SE-20110	US-2014-107993	25.11.2017	Sanjit Engle	Consumer	51,016	7	AZ	21.12.2020	19	21.12.2020	
5	SE-20110	CA-2015-111164	11.04.2018	Sanjit Engle	Consumer	85,14	3	AZ	21.12.2020	19	21.12.2020	
6	SE-20110	CA-2015-111164	11.04.2018	Sanjit Engle	Consumer	21,99	1	AZ	21.12.2020	19	21.12.2020	
7	SE-20110	CA-2015-111164	11.04.2018	Sanjit Engle	Consumer	406,6	5	AZ	21.12.2020	19	21.12.2020	
8	SE-20110	CA-2016-158841	02.02.2019	Sanjit Engle	Consumer	8749,95	5	AZ	21.12.2020	19	21.12.2020	
53	SE-20110	CA-2016-158841	02.02.2019	Sanjit Engle	Consumer	36,4	8	AZ	21.12.2020	19	21.12.2020	

доц. Сахнюк П.А.

# RFM-анализ: применение

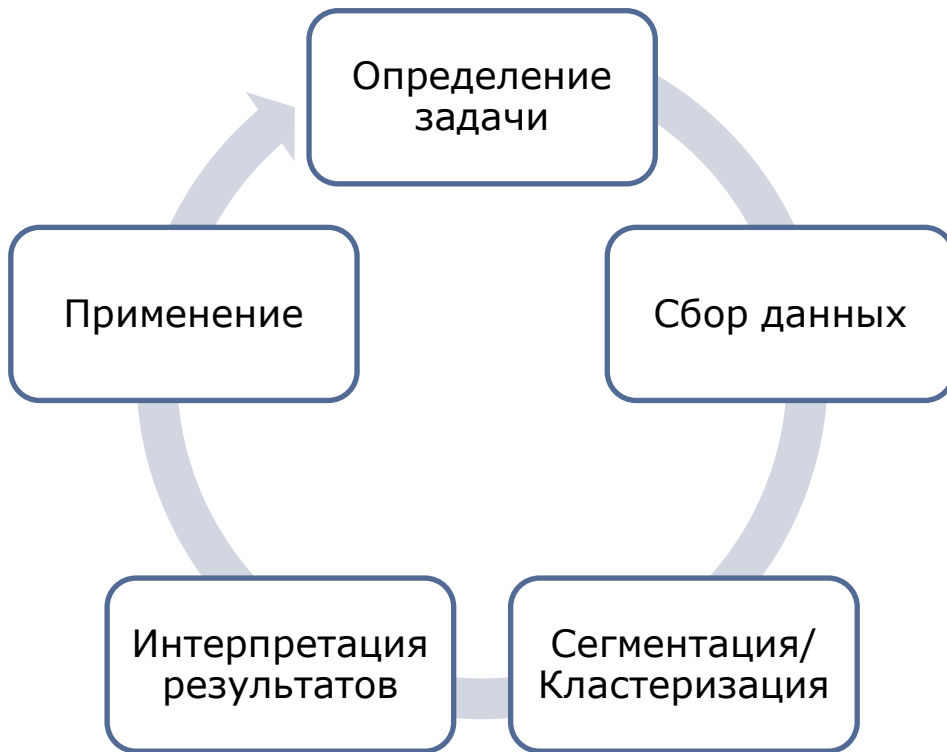
1. Оценка привлекательности клиента
2. Контроль изменений в поведении
3. Оценка миграции между группами
4. Оценка эффективности программы лояльности
5. Выбор стратегии работы с каждым сегментом

# RFM-анализ: выводы

Клиенты	R	F	M	Количество клиентов
Лояльные				
Промежуточные				
Нелояльные				
Потерянные				

# Продвинутые методы

## Цикл работ



доц. Сахнюк П.А.

# Математический аппарат

Разработано множество алгоритмов кластеризации. Некоторые наиболее популярные:

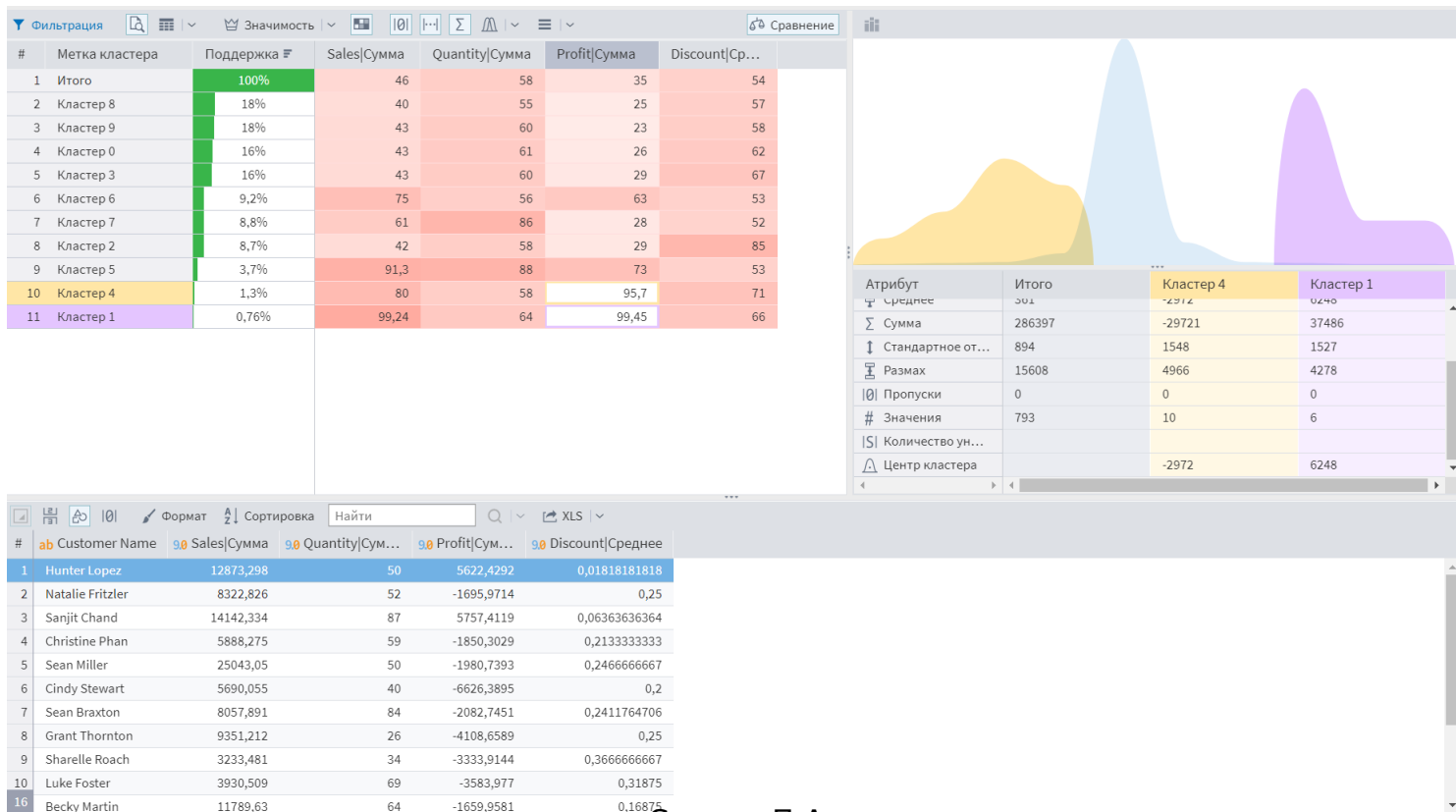
- Карты Кохонена
- К-средних
- ЕМ-кластеризация
- Иерархическая

## Требования к данным

1. Оцифрованы и систематизированы
2. Содержат значимые факторы
3. Очищены от ошибок
4. Рассчитаны агрегаты
5. Сведены в одну таблицу

**Соц-дем**

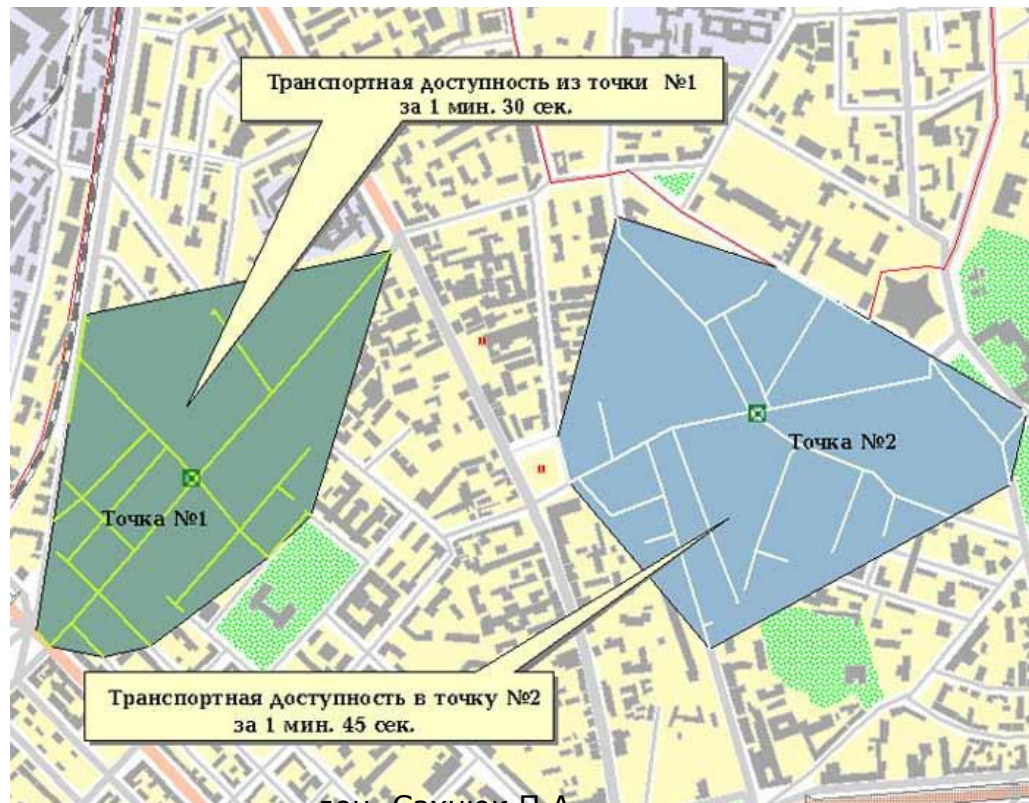
# Сегментация: профили кластеров



доц. Сахнюк П.А.



# Сегментация: география

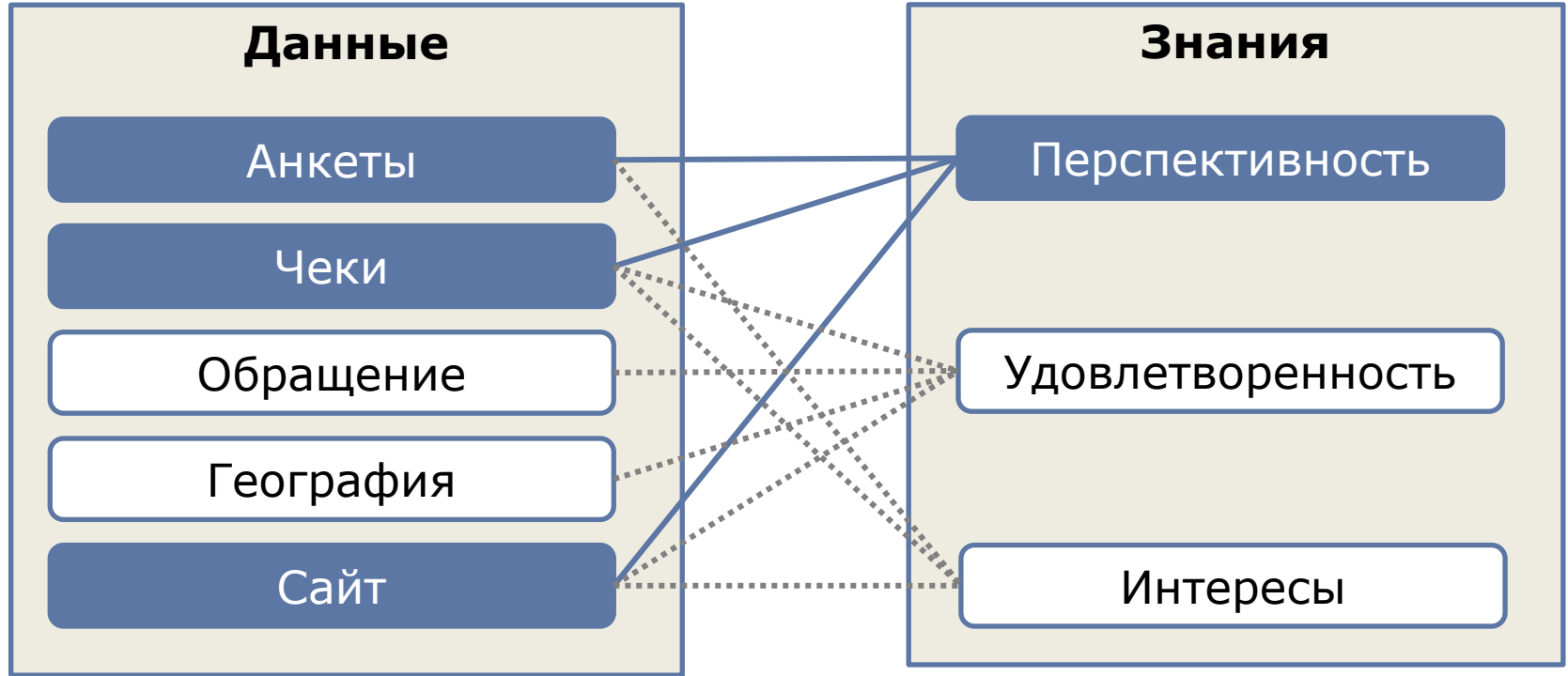


доц. Сахнюк П.А.

## Предпочтения: алгоритм

1. Консолидировать данные о транзакциях (чеках)
2. Определить товары-маркеры
3. Кластеризовать транзакции
4. Рассчитать предпочтения по товарам-маркерам

# Комплексная аналитика



# Оценка сложности

Метод	Сложность	Специалист
ABC	Простой	Маркетолог
XYZ	Простой	Маркетолог
RFM	Простой	Маркетолог
Соц-дем.	Средний	Аналитик
География	Средний	Аналитик
Поведение	Сложный	Data Mining аналитик
Предпочтения	Сложный	Data Mining аналитик
Комплексная	Сложный	Команда аналитиков

# Технологии анализа



доц. Сахнюк П.А.

# XYZ анализ

**XYZ-анализ** — анализ, который позволяет произвести классификацию ресурсов компании в зависимости от характера их потребления и точности прогнозирования изменений в их потребности в течение определенного временного цикла.

Алгоритм проведения можно представить в четырёх этапах:

1. Определение коэффициентов вариации для анализируемых ресурсов;
2. Сортировка ресурсов в соответствии с возрастанием коэффициента вариации;
3. Распределение по категориям X, Y, Z.
4. Графическое представление результатов анализа.

Рассчитывается по формуле

$$V = \frac{\sigma}{\bar{x}}, \quad \sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}, \quad \bar{x} = \frac{\sum_{i=1}^n x_i}{n},$$

где:

$V$  — коэффициент вариации,

$\sigma$  — среднеквадратичное отклонение,

$\bar{x}$  — среднеарифметическое,

$x_i$  —  $i$ -е значение статистического ряда,

$n$  — количество значений в статистическом ряде.

**Категория X** — ресурсы характеризуются стабильной величиной потребления, незначительными колебаниями в их расходе и высокой точностью прогноза. Значение коэффициента вариации находится в интервале от 0 до 10 %.

**Категория Y** — ресурсы характеризуются известными тенденциями определения потребности в них (например, сезонными колебаниями) и средними возможностями их прогнозирования. Значение коэффициента вариации — от 10 до 25 %.

**Категория Z** — потребление ресурсов нерегулярно, какие-либо тенденции отсутствуют, точность прогнозирования невысокая. Значение коэффициента вариации — свыше 25 %.

Реальное значение коэффициента вариации для разных групп может отличаться по следующим причинам:

- сезонность продаж,
- тренд,
- акции,
- дефицит и т. д.

Есть несколько разновидностей XYZ-анализа, например анализ плановых данных с фактическими, что дает более точный % отклонения от прогноза. Очень часто XYZ-анализ проводят совместно с ABC-анализом позволяя выделить более точные группы, относительно их свойств.

```
1 SELECT
2   k.Customer_ID,
3   CASE
4     WHEN k.VAR < 10 and k.VAR <=>0 THEN 'X'
5     WHEN k.VAR >= 10 and k.VAR<25 THEN 'Y'
6   ELSE
7     'Z'
8   END
9   AS XYZ
10  FROM (
11    SELECT
12      Customer_ID,
13      Round(SQRT(STDDEV(Quantity))*100/AVG(Quantity),2) AS VAR,
14    FROM
15      `big-query-exam-283517.Demo1.Orders`
16    GROUP BY
17      Customer_ID) AS k
18  ORDER BY
19    XYZ
20
```

Query results

JOB INFORMATION		RESULTS	JSON	EXECUTION DETAILS
Row	Customer_ID	XYZ		
1	SC-20305	Y		
2	IM-15055	Y		
3	PC-19000	Y		
4	NB-18580	Y		
5	NW-18400	Z		
6	CS-12175	Z		

# ABC анализ

**ABC-анализ** — метод, позволяющий классифицировать ресурсы фирмы по степени их важности. Этот анализ является одним из методов рационализации и может применяться в сфере деятельности любого предприятия. В его основе лежит принцип Парето — 20 % всех товаров дают 80 % оборота. По отношению к ABC-анализу правило Парето может прозвучать так: надёжный контроль 20 % позиций позволяет на 80 % контролировать систему, будь то запасы сырья и комплектующих, либо продуктовый ряд предприятия и т.п.

- ABC-анализ — анализ товарных запасов путём деления на три категории:
- А — наиболее ценные, 20 % — ассортимента (номенклатура); 80 % — продаж
- В — промежуточные, 30 % — ассортимента; 15 % — продаж
- С — наименее ценные, 50 % — ассортимента; 5 % — продаж



# Порядок проведения ABC-анализа

1. Определяем цель анализа (а зачем, собственно, нужен вам этот анализ?).
2. Определяем действия по итогам анализа (что будем делать с полученными результатами?).
3. Выбираем объект анализа (что будем анализировать?) и параметр анализа (по какому признаку будем анализировать?). Обычно объектами ABC анализа являются поставщики, товарные группы, товарные категории, товарные позиции. Каждый из этих объектов имеет разные параметры описания и измерения: объём продаж (в денежном или количественном измерении), доход (в денежном измерении), товарный запас, оборачиваемость и т. д.
4. Составляем рейтинговый список объектов по убыванию значения параметра.
5. Рассчитываем долю параметра от общей суммы параметров с накопительным итогом. Доля с накопительным итогом высчитывается путём прибавления параметра к сумме предыдущих параметров.
6. Выделяем группы А, В и С: присваиваем значения групп выбранным объектам.
  - Вероятности возникновения спроса на материальные ресурсы А, В и С подчинены различным законам. Установлено, что в большинстве промышленных и торговых фирм примерно 80 % стоимости объёма продаж составляют всего около 10 % наименований номенклатуры (группа А), 15 % стоимости — 25 % наименований (группа В), 5 % стоимости — 65 % наименований (группа С). Существует множество способов выделения групп в ABC-анализе.
  - Анализ ABC широко используется при планировании и формировании ассортимента на различных уровнях гибких логистических систем, в производственных системах, системах снабжения и сбыта.

```

10 SELECT
11 DISTINCT m.Customer_ID,
12 m.Sales,
13 SUM(m.Sales) OVER () AS total_sales,
14 m.Sales * 100 / SUM(m.Sales) OVER () AS percent_sales,
15 SUM(m.Sales) OVER (ORDER BY m.Sales DESC) AS cumulative_sum_Customer
16 FROM (
17 SELECT
18 Customer_ID,
19 SUM(Sales) AS Sales,
20 FROM
21 `big-query-exam-283517.Demo1.Orders`
22 GROUP BY
23 Customer_ID) AS m

```

## Query results

JOB INFORMATION		RESULTS	JSON	EXECUTION DETAILS		EXECUTION GRAPH
Row	Customer_ID	Sales	total_sales	percent_sales	cumulative_sum_Customer	
1	SM-20320	25043.07	2297201.07...	1.0901...	25043.07	
2	TC-20980	19052.22	2297201.07...	0.8293...	44095.29	
3	RB-19360	15117.34...	2297201.07...	0.6580...	59212.64	

Оконная функция или аналитическая функция, вычисляет значения для группы строк и возвращает один результат для каждой строки. Это отличается от агрегатной функции, которая возвращает один результат для группы строк.

Оконная функция включает предложение OVER, которое определяет окно строк вокруг оцениваемой строки. Для каждой строки результат оконной функции вычисляется с использованием выбранного окна строк в качестве входных данных, возможно, выполняя агрегирование.

С помощью оконных функций вы можете вычислять скользящие средние значения, ранжировать элементы, вычислять кумулятивные суммы и выполнять другие виды анализа.

```

1 SELECT
2   k.Customer_ID,
3   k.cumulative_sum_Customer *100/k.total_sales AS percent_cumulative_sum,
4   CASE
5     WHEN k.cumulative_sum_Customer*100/k.total_sales < 50 THEN 'A'
6     WHEN k.cumulative_sum_Customer*100/k.total_sales < 80 THEN 'B'
7   ELSE
8     'C'
9   END
10  AS ABC
11  FROM (
12    SELECT
13      DISTINCT m.Customer_ID,
14      m.Sales,
15      SUM(m.Sales) OVER () AS total_sales,
16      m.Sales *100/ SUM(m.Sales) OVER () AS percent_sales,
17      SUM(m.Sales) OVER (ORDER BY m.Sales DESC) AS cumulative_sum_Customer
18    FROM (
19      SELECT
20        Customer_ID,
21        SUM(Sales) AS Sales,
22      FROM
23        `big-query-exam-283517.Demo1.Orders`
24      GROUP BY
25        Customer_ID) AS m ) AS k

```

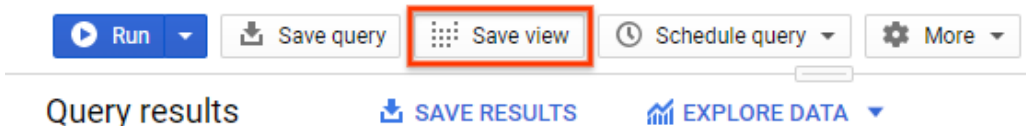
## Query results

JOB INFORMATION		RESULTS	JSON	EXECUTION DETAILS
Row	Customer_ID	percent_cumulative_sum	ABC	
1	SM-20320	1.0901557694294475	A	
2	TC-20980	1.9195224386692455	A	
3	RB-19360	2.5775993565944146	A	
4	TA-21385	3.2129647231968246	A	
5	AB-10105	3.8430171025473188	A	
6	KL-16645	4.4600823731986159	A	
7	SC-20095	5.0757159015253288	A	
8	HL-15040	5.6361065511779529	A	
9	SE-20110	6.1675985550537824	A	
10	CC-12370	6.6955923888717326	A	
11	TS-21370	7.213254954647919	A	
12	GT-14710	7.72779937804922	A	

JOB INFORMATION		RESULTS	JSON	EXECUTION DETAILS
Row	Customer_ID	percent_cumulative_sum	ABC	
168	JF-15565	49.812224316959771	A	
169	GH-14425	49.993873196306673	A	
170	RS-19765	50.17495312240996	B	
171	MY-18295	50.355725282680638	B	
172	PM-19135	50.534885046000788	B	
173	SV-20785	50.713594696349332	B	

JOB INFORMATION		RESULTS	JSON	EXECUTION DETAILS
Row	Customer_ID	percent_cumulative_sum	ABC	
389	LH-16900	79.366076562031211	B	
390	KN-16705	79.465353026324337	B	
391	JF-15190	79.564586394694658	B	
392	FH-14350	79.6634584538131	B	
393	CC-12670	79.762260862955287	B	
394	BE-11410	79.86070413940736	B	
395	KM-16225	79.9591260855542	B	
396	MC-17605	80.057428320804334	C	
397	JH-16180	80.155651764518822	C	
398	MY-17380	80.253783792639453	C	
399	KD-16615	80.351446554045012	C	
400	TS-21505	80.449099303266479	C	

# ABC + XYZ (кросс-анализ)



или на SQL используя: CREATE VIEW

- View (представление) — это виртуальная таблица, определяемая SQL-запросом. После того как вы создадите представление, пользователь сможет запрашивать представление так же, как и таблицу. Результаты запроса содержат только данные из таблиц и полей, указанных в запросе, определяющем представление.
- Запрос, определяющий представление, выполняется каждый раз при запросе представления.

доц. Сахнюк П.А.

```
1 WITH
2   A AS(
3     SELECT
4       *,
5       Customer_ID AS Customer
6     FROM
7       `big-query-exam-283517.Demo.Orders`
8     #GROUP BY 1
9   ),
10  B AS(
11    SELECT
12      Customer_ID,
13      XYZ.XYZ
14    FROM
15      big-query-exam-283517.Demo1.XYZ),
16  C AS(
17    SELECT
18      Customer_ID,
19      ABC.ABC
20    FROM
21      big-query-exam-283517.Demo1.ABC)
22 SELECT
23   *,
24   CONCAT(ABC,XYZ) AS ABC_XYZ
25 FROM
26   A
27 LEFT OUTER JOIN
28   B
29 ON
30   A.Customer=B.Customer_ID
31 LEFT OUTER JOIN
32   C
33 ON
34   A.Customer=C.Customer_ID
```

# ABC + XYZ (кросс-анализ)

	X	Y	Z
A	AX	AY	AZ
B	BX	BY	BZ
C	CX	CY	CZ

Товары обеспечивают основной товарооборот и стабильно продаются, поэтому необходимо обеспечивать постоянное их наличие.

Категория, на которую надо обратить внимание. Важные товары, но с абсолютно не стабильными продажами.

Категория требует вдумчивого анализа. Эти товары можно выводить из ассортимента, если они не новые, элитные и т.п.

	X	Y	Z
AA	AAX	AAZ	AAZ
AB	ABX	ABY	ABZ
BA	BAX	BAY	BAZ
AC	ACX	ACY	ACZ
BB	BBX	BBY	BBZ
CA	CAX	CAY	CAZ
BC	BCX	BCY	BCZ
CB	CBX	CBY	CBZ
CC	CCX	CCY	CCZ

Товары-лидеры

Обратить внимание!

Товары-аутсайдеры

Товары-среднячки

# RFM анализ

RFM-анализ — это анализ клиентов компании с целью их сегментации по ценности для бизнеса. Широко используется в маркетинге, директ-маркетинге и, особенно, в розничной торговле и сфере услуг.

Анализ использует 3 измерения:

**Recency (давность)** — как давно клиент совершил последнюю покупку? Чем меньше времени прошло с момента последней активности клиента, тем больше вероятность, что он повторит действие.

**Frequency (частота)** — как часто клиент совершал покупки? Чем больше покупок совершит клиент, тем больше вероятность того, что он их повторит в будущем.

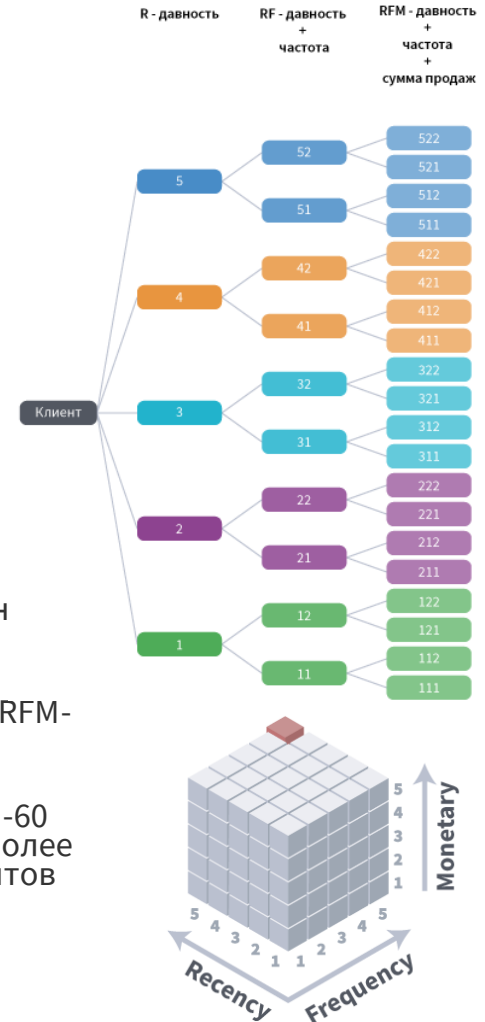
**Monetary (деньги)** — сколько клиент тратит? Чем больше денег было потрачено, тем больше вероятность того, что клиент будет совершать покупки и в дальнейшем.

Предполагается, что клиент, проявивший себя недавно, показывающий повышенную активность и тратящий на покупки больше денег, будет проявлять активное потребительское поведение и в дальнейшем.

История покупок клиента или продаж товара представляется в виде таблицы или RFM-матрицы, в которой три колонки – давность, частота и деньги. Каждая колонка разбивается на категории.

Например, давность можно разбить на интервалы 1-30 дней (настоящее время), 31-60 дней (недавно), 61-90 дней (давно). Частоту покупок можно разделить на частые (более 10 в месяц), редкие (3-10 в месяц) и разовые (менее 3-х раз в месяц). Затраты клиентов можно разделить на высокие (5-10 тыс.), средние (2-5 тыс.) и низкие (менее 2 тыс.). Тогда можно составить RFM-матрицу вида:

доц. Сахнюк П.А.



Наиболее «перспективные» клиенты окажутся в верхнем левом углу таблицы, в выделенном сегменте (содержит клиентов, которые делают покупки часто, в настоящее время или недавно, тратя при этом средние или большие суммы). Наименее «перспективные» окажутся в нижнем выделенном сегменте (разовые покупки, сделанные давно или недавно на низкие и средние суммы).

В зависимости от особенностей анализа могут использоваться и другие представления RFM-матрицы. Например, иногда используют не 3 категории для каждого измерения, когда получается 27 сегментов, а 5 категорий, что позволяет создать 125 сегментов. В этом случае результаты анализа оказываются более детальными.

Иногда сегментам матрицы присваивают баллы от 1 (самые «непривлекательные») до 5 (самый «привлекательный» сегмент). В этом случае «лучший» сегмент будет обозначен 5R-5F-5M, а «худший» 1R-1F-1M. Тогда клиентов, попавших в сегмент 3R-3F-3M можно интерпретировать как средних по привлекательности.

Преимуществом метода является простота (не требует специального статистического ПО), а результаты легко интерпретируются. При использовании в директ-маркетинге применение анализа RFM может увеличить количество откликов на рекламные акции.

Частота	Давность	Деньги		
		Высокие	Средние	Низкие
Часто	Настоящее время			
	Недавно			
	Давно			
Редко	Настоящее время			
	Недавно			
	Давно			
Разово	Настоящее время			
	Недавно			
	Давно			

Существуют несколько модификаций RFM-анализа:

RFD – Recency (давность), Frequency (частота), Duration (продолжительность) — модифицированная версия RFM-анализа для изучения поведения потребителей бизнес-продуктов, ориентированных на зрителей, читателей, интернет-сёрферов (например, количество времени, проведенного серферами в онлайн-кинотеатре).

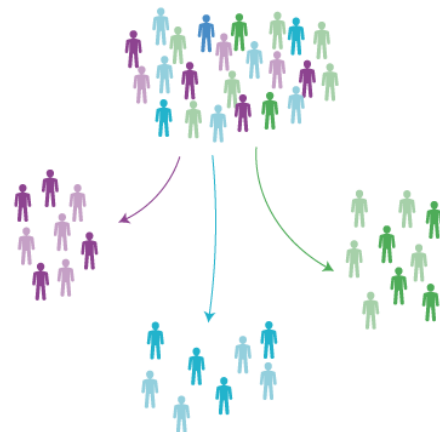
RFE — Recency (давность), Frequency (частота), Engagement (вовлечённость) расширенная версия RFD-анализа, где можно оценить вовлечение клиента в работу с бизнес-продуктом. Кроме продолжительности при этом используют, например, количество посещённых web-страниц, количество переходов по ссылкам и другие действия клиента, свидетельствующие о его активном пользовании бизнес-продуктом. Предполагается, что наиболее активные клиенты более склонны к отклику на маркетинговые акции.

RFM-I – Recency (давность), Frequency (частота), Monetary – Interactions (деньги-взаимодействие) — является версией RFM-анализа для учета давности и частоты маркетинговых взаимодействий с клиентом, например, для изучения возможных сдерживающих эффектов частых рекламных мероприятий (когда слишком назойливая реклама отталкивает клиента).

```

1 WITH
2   raw_RFM AS (
3     SELECT
4       Customer_ID,
5       Recency,
6       NTILE(5) OVER (ORDER BY Recency DESC) AS r,
7       NTILE(5) OVER (ORDER BY Frequency ASC) AS f,
8       NTILE(5) OVER (ORDER BY Monetary ASC) AS m,
9     FROM (
10      SELECT
11        Customer_ID,
12        DATE_DIFF(CAST(DATE '2020-12-30' AS DATE), CAST(MAX(Order_Date) AS DATE), DAY) AS Recency,
13        SUM(Sales) AS Monetary,
14        Count (DISTINCT Order_ID) AS Frequency
15      FROM
16        `big-query-exam-283517.Demo1.Orders`
17      GROUP BY
18        Customer_ID ) )
19 SELECT
20   *, concat(r,f,m) as RFM,
21   CASE
22     WHEN r>=4 and f>=4 and m>=4 THEN 'champions'
23     WHEN r>=3 and f>=3 and m>=3 THEN 'loyals'
24     ELSE
25       'Risk'
26   END
27   AS RFM_segmentation
28 FROM
29   raw_RFM

```



## Query results

JOB INFORMATION		RESULTS	JSON	EXECUTION DETAILS			EXECUTION GRAPH
Row	Customer_ID	Recency	r	f	m	RFM	RFM_segmentation
1	MG-18205	264	1	1	1	111	Risk
2	RB-19645	649	1	2	1	121	Risk
3	AG-10300	118	2	2	1	221	Risk
4	SF-20065	59	3	4	1	341	Risk
5	MK-18160	240	1	1	2	112	Risk
6	PJ-18835	39	4	4	2	442	Risk
7	SS-20410	110	2	3	4	234	Risk
8	GM-14680	152	2	3	4	234	Risk
9	PO-19180	160	2	5	5	255	Risk
10	MW-18220	17	5	3	5	535	loyals