**Самостоятельная работа 5:**

**Использование инструментов "Prediction Calculator" и "ShoppingbasketAnalysis"**

**Аннотация:**Лабораторная работа посвящена использованию инструментов "Расчет прогноза" ("PredictionCalculator") и "Анализ покупательской корзины" ("ShoppingBasketAnalysis").

**Ключевые слова:**[SQL](https://www.intuit.ru/studies/courses/2312/612/lecture/13268?page=1#keyword1), [server](https://www.intuit.ru/studies/courses/2312/612/lecture/13268?page=1#keyword2), [excel](https://www.intuit.ru/studies/courses/2312/612/lecture/13268?page=1#keyword3), [значение](https://www.intuit.ru/studies/courses/2312/612/lecture/13268?page=1#keyword4), [интервал](https://www.intuit.ru/studies/courses/2312/612/lecture/13268?page=1#keyword6), [диапазон](https://www.intuit.ru/studies/courses/2312/612/lecture/13268?page=1#keyword10), [ПО](https://www.intuit.ru/studies/courses/2312/612/lecture/13268?page=1#keyword11), [таблица](https://www.intuit.ru/studies/courses/2312/612/lecture/13268?page=1#keyword13), [лист](https://www.intuit.ru/studies/courses/2312/612/lecture/13268?page=1#keyword14), [истина](https://www.intuit.ru/studies/courses/2312/612/lecture/13268?page=1#keyword20), [печать](https://www.intuit.ru/studies/courses/2312/612/lecture/13268?page=1#keyword23), [прибыль](https://www.intuit.ru/studies/courses/2312/612/lecture/13268?page=1#keyword26), [максимум](https://www.intuit.ru/studies/courses/2312/612/lecture/13268?page=1#keyword28), [график](https://www.intuit.ru/studies/courses/2312/612/lecture/13268?page=1#keyword33), [продажа](https://www.intuit.ru/studies/courses/2312/612/lecture/13268?page=1#keyword35), [затраты](https://www.intuit.ru/studies/courses/2312/612/lecture/13268?page=1#keyword37), [алгоритм](https://www.intuit.ru/studies/courses/2312/612/lecture/13268?page=2#keyword41), [меню](https://www.intuit.ru/studies/courses/2312/612/lecture/13268?page=2#keyword42), [идентификатор](https://www.intuit.ru/studies/courses/2312/612/lecture/13268?page=2#keyword45), [анализ](https://www.intuit.ru/studies/courses/2312/612/lecture/13268?page=2#keyword46), [стоимость](https://www.intuit.ru/studies/courses/2312/612/lecture/13268?page=2#keyword47)

### Расчет прогноза

Инструмент **Prediction Calculator** помогает сгенерировать и настроить "калькулятор", который позволяет оценить шансы на получение ожидаемого значения целевого параметра без подключения к аналитическим службам *SQL* *Server*. В частности, такая возможность может быть очень полезна для удаленных пользователей.

В качестве учебного набора данных в этой части лабораторной будем использовать локализованный пример для *Excel*, взятый с <http://russiandmaddins.codeplex.com/>

Перейдем на набор данных **"Клиенты"** и на вкладке **Analyze** выберем **Prediction Calculator**. В окне настроек надо указать целевой столбец и искомое *значение* ([рис. 8.1](https://www.intuit.ru/studies/courses/2312/612/lecture/13268?page=1#image.8.1)). Если значения целевого столбца рассматриваются как числовые из непрерывного диапазона, то можно указать, как точное *значение*, так и желаемый *интервал*. В противном случае - только точное *значение*.



**Рис. 8.1.**Настройки инструмента Prediction Calculator

Пусть цель анализа - определить, купит ли клиент велосипед. В качестве целевого столбца указываем "Приобрел велосипед" и *значение* "Да". Далее можно указать столбцы для анализа. Как и ранее, рекомендуется исключать из рассмотрения столбцы с уникальными значениями и столбцы, один из которых дублирует другой (например, точное *значение* заработной платы и *диапазон*заработной платы).

Инструмент всегда формирует отчет **Prediction Calculator Report**, кроме того *по* умолчанию формируются два необязательных отчета - **Prediction Calculator** ("калькулятор" прогноза в виде таблицы *Excel*) и **Printable Calculator** (*таблица* калькулятора для печати и ручной обработки).

Чтобы лучше разобраться с результатами работы инструмента, перейдем сначала на *лист* с отчетом **Prediction Calculator**.В верхней части отчета расположен сам калькулятор ([рис. 8.2](https://www.intuit.ru/studies/courses/2312/612/lecture/13268?page=1#image.8.2)), в нижней - *таблица* баллов, соответствующих различным значениям параметров ([рис. 8.3](https://www.intuit.ru/studies/courses/2312/612/lecture/13268?page=1#image.8.3)).

Работая с калькулятором, можно описать анализируемый пример, указывая значения для каждого параметра. Значения в столбец **Value** можно вводить или выбирать из выпадающих списков (что лучше, т.к. меньше шансов ввести некорректное *значение* или *диапазон*). Для описываемого примера рассчитывается сумма баллов, которая сравнивается с рекомендуемым пороговым значением. Если *значение* выше "порога", то прогноз получает *значение* "*истина*" (на рисунке сумма баллов 572, пороговое *значение* 565). Вторая часть отчета поясняет полученный результат, показывая, сколько баллов за какое *значение* ставится.



[увеличить изображение](https://www.intuit.ru/EDI/03_01_17_9/1483395677-6027/tutorial/1036/objects/8/files/21_02.jpg)
**Рис. 8.2.**"Калькулятор"



[увеличить изображение](https://www.intuit.ru/EDI/03_01_17_9/1483395677-6027/tutorial/1036/objects/8/files/21_03.jpg)
**Рис. 8.3.**Таблица баллов для параметров

Представленный на [рис. 8.4](https://www.intuit.ru/studies/courses/2312/612/lecture/13268?page=1#image.8.4) отчет **"Printable Calculator"** позволяет вывести на *печать* готовую форму для ручного подсчета баллов и получения оценки без использования компьютера. Это может быть удобно, например, для торговых представителей или других сотрудников, работающих вне офиса и не имеющих доступа к компьютеру. Все что нужно для расчета прогноза - отметить варианты, просуммировать баллы и сравнить с пороговым значением.

Теперь перейдем к более интересному вопросу - как же было определено пороговое *значение*. Отчет **Prediction Calculator Report**позволяет с этим разобраться ([рис. 8.5](https://www.intuit.ru/studies/courses/2312/612/lecture/13268?page=1#image.8.5)). *По* итогам анализа формируется прогноз, который может быть отнесен к одной из четырех категорий [[1](https://www.intuit.ru/studies/courses/2312/612/literature#literature.1)]:

* **истинный позитивный прогноз (TruePositive)** - верный прогноз. Например, клиент, для которого прогноз показал истину, на самом деле заинтересован в покупке велосипеда. Магазин получил прибыль;
* **истинный негативный прогноз (TrueNegative)** - верный негативный прогноз. Клиент, для которого прогноз показал незаинтересованность в покупке, на самом деле не собирается покупать велосипед. Магазин не получил прибыли, но и не понес затрат (на рассылку рекламных предложений и проч.);
* **ложный позитивный прогноз (FalsePositive; ошибка 1 рода)** - неверный прогноз, показывающий,что клиент хочет сделать покупку, хотя на самом деле это не так (может привести магазин к затратам на сопровождение клиента);
* **ложный негативный прогноз (FalseNegative; ошибка 2 рода)** - неверный прогноз, показывающий, что клиент не хочет сделать покупку, хотя на самом деле он в ней заинтересован (может привести к упущенной прибыли).



[увеличить изображение](https://www.intuit.ru/EDI/03_01_17_9/1483395677-6027/tutorial/1036/objects/8/files/21_04.jpg)
**Рис. 8.4.**Отчет "PrintableCalculator"

Отчет **Prediction Calculator Report** позволяет указать *прибыль* от истинных прогнозов и убыток от ложных. На основе этих данных определяется пороговое *значение*, обеспечивающее *максимум* прибыли. *По* умолчанию, для истинного позитивного прогноза указывается *прибыль* 10 (долларов или других единиц), для ложного позитивного - такой же убыток ([рис. 8.5](https://www.intuit.ru/studies/courses/2312/612/lecture/13268?page=1#image.8.5), *таблица* в левой верхней части экрана). В этом случае *максимум* прибыли (*график* на [рис. 8.5](https://www.intuit.ru/studies/courses/2312/612/lecture/13268?page=1#image.8.5) справа вверху) как раз и будет соответствовать пороговому значению для прогноза в 565 баллов.



[увеличить изображение](https://www.intuit.ru/EDI/03_01_17_9/1483395677-6027/tutorial/1036/objects/8/files/21_05.jpg)
**Рис. 8.5.**Отчет Prediction Calculator Report

В нижней части отчета **Prediction Calculator** слева располагается *таблица* с относительными весами значений рассматриваемых параметров (ее мы уже встречали в таблице **Prediction Calculator**) и графиком потерь от ложных прогнозов.

Пусть *продажа* велосипеда приносит магазину не 10, а 50 долларов. В этом случае, *прибыль* от одной продажи будет перекрывать *затраты* на сопровождение до 5 отказавшихся от покупки клиентов. Соответственно изменится и соотношение прибыли/*затраты*. На [рис. 8.6](https://www.intuit.ru/studies/courses/2312/612/lecture/13268?page=1#image.8.6) показано, что в этом случае, для максимизации прибыли рекомендуется установить пороговое *значение* для прогноза в 443 балла. Новое *значение* будет автоматически подставлено и в таблицу **Prediction Calculator**.

**Задание.** Проведите анализ для двух различных наборов значений прибыли от истинных прогнозов и убытков от ложных. Прокомментируйте результаты.



[увеличить изображение](https://www.intuit.ru/EDI/03_01_17_9/1483395677-6027/tutorial/1036/objects/8/files/21_06.jpg)
**Рис. 8.6.**Отчет Prediction Calculator Report: при вводе новой оценки прибыли от правильного прогноза меняется рекомендуемое пороговое значение

### Анализ покупательской корзины

В наборе **Table Analysis Tools** нам осталось рассмотреть инструмент **Shopping Basket Analysis**. Он позволяет, например, на основе данных о покупках выделить товары, чаще всего встречающиеся в одном заказе, и сформировать рекомендации относительно совместных продаж.

В процессе анализа используется *алгоритм* **MicrosoftAssociationRules**.

Для изучения этого инструмента, вместо использованного ранее локализованного набора данных, обратимся к примеру из поставки надстроек интеллектуального анализа (в предыдущем файле нужного набора данных просто нет). Через *меню* **"Пуск"** найдите **"Надстройки интеллектуального анализа данных" -> "Образцы данных Excel"**. В этой книге *Excel* c первого листа ([рис. 8.7](https://www.intuit.ru/studies/courses/2312/612/lecture/13268?page=2#image.8.7)) перейдите *по* ссылке **"Поиск взаимосвязей и покупательское поведение"**. Соответствующий набор данных ([рис. 8.8](https://www.intuit.ru/studies/courses/2312/612/lecture/13268?page=2#image.8.8)) содержит информацию о заказах (номер заказа - **Order Number**), включенных в них товарах (их категории - **Category** и собственно товаре - **Product**) и ценах.



[увеличить изображение](https://www.intuit.ru/EDI/03_01_17_9/1483395677-6027/tutorial/1036/objects/8/files/21_07.jpg)
**Рис. 8.7.**Образцы данных



[увеличить изображение](https://www.intuit.ru/EDI/03_01_17_9/1483395677-6027/tutorial/1036/objects/8/files/21_08.jpg)
**Рис. 8.8.**Запуск инструмента Shopping Basket Analysis

Проанализируем, какие категории чаще всего попадают в один заказ. Запустим инструмент **Shopping Basket Analysis**. В его настройках надо указать *идентификатор* транзакции (**TransactionID**), в нашем случае, это **Order Number** и предмет анализа (мы будем проводить *анализ* для категорий - **Category**). Необязательным параметром, количественно характеризующим предмет анализа (**Item Value**), в нашем случае будет цена. Если **Item Value** не указан, то анализироваться будет только частота выявленных сочетаний.

Результаты работы **Shopping Basket Analysis** отображаются в двух отчетах - **Bundled Items** и **Recommendations**.



[увеличить изображение](https://www.intuit.ru/EDI/03_01_17_9/1483395677-6027/tutorial/1036/objects/8/files/21_09.jpg)
**Рис. 8.9.**Отчет Shopping Basket Bundled Items



[увеличить изображение](https://www.intuit.ru/EDI/03_01_17_9/1483395677-6027/tutorial/1036/objects/8/files/21_10.jpg)
**Рис. 8.10.**Отчет Shopping Basket Recommendations

Первый из отчетов содержит информацию о наиболее часто встречающихся в "одном чеке" сочетаниях категорий товаров. Так, например, в первой строке отчета на [рис. 8.9](https://www.intuit.ru/studies/courses/2312/612/lecture/13268?page=2#image.8.9) мы видим, что чаще всего встречается сочетание категорий "дорожные велосипеды" и "шлемы" (RoadBikes, Helmets). В рассматриваемом наборе оно выявлено в 805 заказах. Дальше указывается средняя цена набора и суммарная *стоимость* всех подобных наборов. Можно сказать, что этот отчет описывает покупательские шаблоны клиентов.

Второй отчет **Shopping Basket Recommendations** содержит рекомендации о товарах, которые могут быть предложены вместе. Например, третья строчка отчета указывает, что людям купившим шлем, стоит также предложить приобрести шины. Это заключение базируется на том, что среди 3794 покупок включающих шлемы, 1617 включали и шины. Доля таких связанных продаж равна 42,62%. Далее приводится средний доход от связанных продаж (общая *стоимость*, деленная на число транзакций, которые содержат "рекомендующий" продукт, в нашем случае - шлем) и общая сумма связанных продаж. Основываясь на подобном отчете, владелец магазина может решить, как разместить товары, какие связанные предложения можно сформировать и т.д.

Для удаления результатов работы инструмента достаточно удалить сформированные отчеты.

**Задание 1.** Проведите анализ аналогичный описанному выше.

**Задание 2.** Проанализируйте, какие товары (а не категории товаров, как было раньше), приобретаются вместе. Опишите полученные результаты.