

Введение

Мы не будем приводить характеристики нашего региона из Википедии, а просто скажем, что республика Башкортостан – это один из ключевых экономических регионов России, благодаря своим природным богатствам и развитой промышленности.

Наш проект посвящен профессиональному спорту. Команды из Уфы представлены в высшей лиге двух самых популярных видах спорта в России – футболе и хоккее. Если о хоккейном клубе «Салават Юлаев» слышали многие, так как он дважды за последние десять лет выигрывал Континентальную хоккейную лигу, то о состоянии дел «ФК Уфа» едва ли знают не заядлые футбольные любители за пределами региона. Так вот, уже два сезона «ФК Уфа» выступает в Российской футбольной Премьер-лиге, и каждый раз чудом умудряется сохранить прописку в турнире (команды, которые занимают два последних места отправляются в нижний дивизион).

Нам стало интересно, а как бы выступила наша команда на полях топ Европейских чемпионатов – Англии, Франции, Испании, Италии и Германии. Быть может причина наших не очень успешных выступлений в том, что конкуренция нашего чемпионата зашкаливает, и нам было бы приятнее болеть за Уфу в Испании, например.

Кроме того, интересно стало посмотреть – кто они болельщики «ФК Уфа». В частности, есть ли среди них люди не из Башкирии, если есть, то сколько (по крайней мере среди пользователей соцсетей)? Какой город в республике самый патриотично-футбольный (болеют за «Уфу», а не за «Зенит», например)? Какая доля из болельщиков интересуется исключительно футболом, несмотря на то, что хоккейная команда куда успешнее и ее популярность заметно выше?

Задачи

Для решения поставленных вопросов были поставлены следующие задачи:

- 1) Написать функцию, оценивающую результат команды по введенным характеристикам (регрессия), и оценить результат нашей команды в топ-5 Европейских чемпионатов
- 2) Проанализировать участников группы «ФК Уфа» с помощью VK API (их личная информация)
- 3) Найти количество участников группы, состоящих в сообществе другой популярной в России футбольной команды (Спартак, Зенит, ЦСКА)
- 4) Найти количество участников группы, состоящих в сообществе хоккейной команды «Салават Юлаев»

Для ответа на поставленные вопросы использовали следующие инструменты:

- VBA
- VK API
- iPython (Jupyter)

Мы не стали использовать VBA для решения абсолютно всех наших задач. В случаях, когда более удобным инструментом оказывались, например, встроенные функции Excel, мы пользовались ими. Использовать VBA ради VBA мы не стали, однако, как оказалось, зачастую просто нет альтернативы, и этот инструмент действительно очень удобен.

Решение

Задача 1. Предсказание места «ФК Уфа» в топ чемпионатах Европы

Пожалуй, самый интересный для нас вопрос в данном проекте.

Мы решили обучить регрессию, целевой переменной которой стало количество очков, набранное футбольными командами за сезон 2015-2016. Признаками стали 17 переменных:

- 11 – характеристики игроков основного состава
- 1 – сыгранность (средняя продолжительность пребывания в команде по всем игрокам)
- 5 – Дамми переменные принадлежности к чемпионату

Как оценить характеристики игроков? Мы скачали данные из футбольного симулятора Fifa 16, в котором каждый игрок оценивается по 100 – балльной шкале. Понятно, что это не самый точный метод оценивания, однако это лучшее из того, что доступно всем. Данный футбольный симулятор является самым популярным в мире и все цифры в нем согласуются с FIFA. Оценивание игроков производится по более чем 20-ти характеристикам, из которых затем выводится общий показатель.

Итак, как мы получили характеристики игроков 98 команд:

Сначала скопировали все ссылки на команды из топ дивизионов Европы. Проблему с лишними ссылками решили с помощью разделителя из «текста по столбцам» в разделе «Данные» (так как все ненужные ссылки содержали символ «?»). Затем просто удалили ненужные строки фильтром.

Макрос(был предложен на лекции):

```
Sub ExtractHL()  
Dim HL As Hyperlink  
i = 1  
For Each HL In Sheets("All").Hyperlinks  
    Sheets("URL").Cells(i, 2) = HL.Address  
    i = i + 1  
Next  
End Sub
```

Затем, воспользовавшись алгоритмом для скачивания новостей, мы скачали составы команд с характеристиками игроков по ссылкам. Удалив с помощью фильтра резервных игроков и запасных, получили данные по основному составу для 98 команд (лист «Состав» в книге Excel). Мы решили исследовать команды не по отдельным характеристикам каждого из игроков (удар, передача, дриблинг и т.д.), а анализируя их суммарные характеристики (в нашей книге это OVA).

Теперь нужно было выписать в строку составы каждой из команд, то есть получить таблицу 98 на 11.

Вручную выделять и транспонировать нам было лень, поэтому написали простой макрос:

```
Sub more11()  
k = 20  
l = 0  
For i = 1 To 98  
    For j = 1 To 11  
        u = l + j + 1  
        Cells(i, k) = Cells(u, 5)  
        k = k + 1  
    Next j
```

```

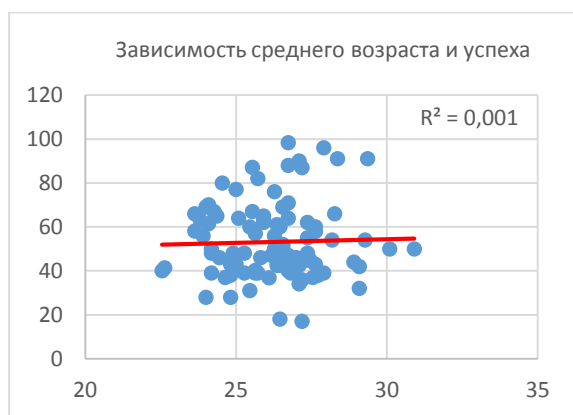
k = 20
l = l + 11
Next i
End Sub

```

На листе «Состав» чуть правее таблицы с командами можно увидеть результат работы макроса. Все получилось, как нам нужно. Далее мы скопировали данный результат на лист «Данные» и добавили названия команд с их количеством набранных за сезон очков. Кроме того, понятно, что чемпионаты могут различаться по уровню, поэтому мы ввели *Дамми-переменную*, характеризующую чемпионат (1 – если клуб принадлежит данной стране, 0 – если не принадлежит). Таким образом, мы ввели 5 Дамми переменных [Англия, Франция, Италия, Испания, Германия]).

Далее из-за малости выборки пришлось сразу подредактировать выбросы. Если состав команды ключевым образом не изменился в сравнении с предыдущим годом, а результат изменился очень сильно, то мы брали результат на основании предыдущих лет. Таким образом, был изменен результат 6 команд (в листе «Данные» они выделены желтым).

Далее с помощью стандартных сводных таблиц мы решили проверить влияние еще двух факторов: среднего возраста команды и сыгранности (как среднее пребывание игроков в команде). Корреляции получились следующие:



Получается, что возраст не сильно влияет на успех команды, а вот количество лет, сыгранное игроками вместе действительно важно.

Поэтому добавим данный фактор в «Данные». Сводная таблица расставила команды по алфавиту, поэтому в данном случае удобно использовать VBA для поиска соответствия команды из листа Corr с листом Данные и последующим экспортом значения:

```

Sub ss1()
For i = 2 To 99
For j = 1 To 98
If Worksheets("Данные").Cells(i, 1) = Worksheets("Corr").Cells(j, 1) Then
Worksheets("Данные").Cells(i, 13) = Worksheets("Corr").Cells(j, 3)
End If
Next j
Next i
End Sub

```

Далее преобразуем подготовленный и удобный для анализа лист «Данные» в формат .CSV, и продолжаем анализ уже в Jupyter Notebook. Продолжение в html документе.

