



**Полгода стажерства  
Дубенюк Анны**

# Самые интересные задачи

## Автоматическая генерация резюме

На пилотной версии портала для сотрудников (тогда еще sberout, теперь nextsteps) резюме выкладывались в формате pdf. Нашей команде резюме присылали в формате excel, где каждая строка представляла собой резюме одного человека. Из каждой строчки методологи руками

С нуля мною была написана программа для генерации pdf-файлов с корректным форматированием. Для использования нужно положить первичный excel-файл в конкретную папку, запустить скрипт, подождать 1-2 минуты (в зависимости от количества строк), в отдельной папке получить готовые pdf-файлы.

## Ликбез по математической статистике

Мною был проведен краткий ликбез по математической статистике для членов команды. Это был интересный преподавательский опыт.

## Разработка системы мониторинга call-центра

Сбор данных для выходного интервью осуществляется с помощью сил call-центра в ПЦП. Помимо получаемых данных выходного интервью необходимо мониторить работу операторов, а также следить за показателями самих кампаний. Из системы телефонии можно выгружать только ежедневные отчеты – по кодам завершения и по производительности. Это неудобно – их два, они ежедневные, с огромным количеством ненужных параметров, непригодные для обработки в excel.

Проанализировав системные файлы, мною были выявлены наиболее значимые метрики и написана программа для автоматического подсчета этих показателей, генерации удобного excel-файла для детального анализа, а также фотографий с основными агрегированными показателями, которые автоматически выгружаются в презентацию. Ежедневно мой скрипт используется для мониторинга работы call-центра, принятия управленческих решений.

# Самые интересные задачи

## Автоматизация создания отчетов по результатам Exit-интервью, полный цикл отчетности

- Раньше каждый месяц методологи делали в ручную презентации с основными результатами выходных интервью. Однотипные презентации в разных разрезах, все подсчеты были в excel. При изменении скрипта возникало много сложностей.
- Мною с нуля была написана программа для генерации отчетов. На вход подаются первичные данные, они автоматически очищаются, группируются в нужных разрезах, подсчитываются показатели с учетом всех особенностей скрипта и генерируются фотографии с визуализацией каждого показателя.
- Далее эти фотографии с помощью другого моего скрипта автоматически выгружаются на шаблоны. Этот механизм работает с помощью составленных таблиц соответствий (фото, слайд, placeholder). Причем скрипт выгрузки написан так, чтобы в дальнейшем можно было выгружать не только фотографии, но и таблицы.
- Нужные разрезы, шаблоны презентаций, цвета для картинок, а также таблицы соответствий для выгрузки на слайды легко меняются, так как хранятся в отдельных файлах.
- Изначально выгрузка происходила с помощью моего кода на VBA, так как подразумевалось, что эта функциональность будет у методолога, однако потом я переписала скрипт выгрузки на python.
- С течением времени список показателей расширялся, появилась возможность динамики. Ежемесячно скрипты применяются для формирования отчетности. Остается в ручную дописать выводы и добавить цитаты из интервью.
- Раньше отчеты отдавались методологу для написания выводов, теперь полный отчетный цикл с момента получения первичных данных осуществляется мной.
- Помимо ежемесячных отчетов с помощью скрипта формируются отдельные отчеты по запросу контрагентов.

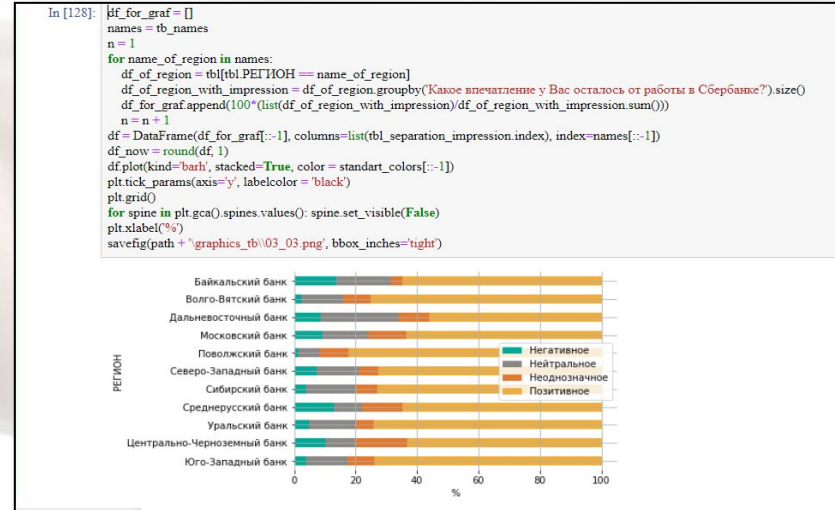


# Автоматизация отчетности

## Первичные данные выходного

СТАЖ	ТЕЛЕФОН	ДАТА УВОЛЬНЕНИЯ	ДИСЦИПЛИНА	ПОДРАЗДЕЛ	РЕГИОН	СТАТЬЯ УИ CALL_ВАС	Правила Уточните, Удобно ли Какое впечатление у Вас осталось от работы в Сбербанке?
1,9	9,81E+09	23.05.2019	Менеджер	Сеть прод Центрально-Черноземный бан	ДА	ДА	Неоднозначное
5,3	9,16E+09	31.05.2019	Старший	Сеть прод Среднерусский банк	ДА	ДА	Позитивное
3,2	9,07E+09	28.05.2019	Мобильн	Сеть прод Московский банк	ДА	ДА	Позитивное
2,9	9,02E+09	31.05.2019	Специали	Сервисы ДСЦ	ДА	ДА	Нейтральное
2,7	9,52E+09	29.05.2019	Инкасатор	Сервисы Сибирский банк	ДА	ДА	Позитивное
5,2	9,09E+09	31.05.2019	Менеджер	Сеть прод Северо-Западный банк	ДА	ДА	Негативное
2,2	9,27E+09	31.05.2019	Менеджер	Технолог ОЦ	ДА	ДА	Позитивное
5	9,64E+09	31.05.2019	Водитель	Сервисы Сибирский банк	ДА	ДА	Нейтральное
2,9	9,26E+09	31.05.2019	Руководит	Розничны ПАО «Сбербанк России» (ЦА)	ДА	ДА	Позитивное
0,7	9,54E+09	27.05.2019	Консультант	Сеть прод Сибирский банк	ДА	ДА	Позитивное
0,6	9,2E+09	31.05.2019	Главный	Технолог ПАО «Сбербанк России» (ЦА)	ДА	ДА	Позитивное
1,6	9,81E+09	28.05.2019	Менеджер	Сеть прод Северо-Западный банк	ДА	ДА	Позитивное
5,2	9,14E+09	30.05.2019	Старший	Сеть прод Байкальский банк	ДА	ДА	Позитивное
0,6	9,14E+09	28.05.2019	Старший	Технолог МЦТП	ДА	ДА	Позитивное
3,2	9,38E+09	29.05.2019	Сервис-м	Сервисы Поволжский банк	ДА	ДА	Позитивное
0,5	9,28E+09	31.05.2019	Консультант	Сеть прод Волго-Ватский банк	ДА	ДА	Позитивное
4,6	9,82E+09	27.05.2019	Ст.менед	Розничны Северо-Западный банк	ДА	ДА	Позитивное
0,3	9,22E+09	30.05.2019	Специали	Сервисы Уральский банк	ДА	ДА	Нейтральное
4,5	9,13E+09	27.05.2019	Руководит	Сеть прод Северо-Западный банк	ДА	ДА	Позитивное
0,2	9,14E+09	29.05.2019	Старший	Сеть прод Сибирский банк	ДА	ДА	Позитивное
2,2	9,12E+09	27.05.2019	Клиентск	Сеть прод Северо-Западный банк	ДА	ДА	Позитивное

## Анализ данных, генерация



## Загрузка фотографий на шаблоны

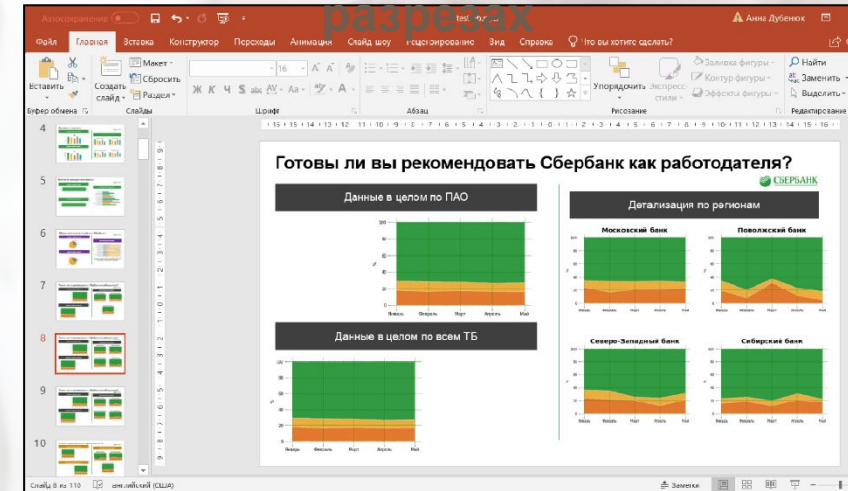
```
In [12]: location = pd.read_csv('photos_tb_new.csv', sep=';', low_memory=False)
location.head()

Out[12]:
```

number	path	slide_of_photo	position_of_photo
0	C:\Users\Anyai\Desktop\study_work\sbergraphics...	2	2
1	C:\Users\Anyai\Desktop\study_work\sbergraphics...	2	1
2	C:\Users\Anyai\Desktop\study_work\sbergraphics...	2	3
3	C:\Users\Anyai\Desktop\study_work\sbergraphics...	2	4
4	C:\Users\Anyai\Desktop\study_work\sbergraphics...	2	5

```
In [13]: pr = pptx.Presentation('format_2019_tb_n.pptx')
for i in range(len(location)):
    try:
        slide = pr.slides[location.iloc[i]['slide_of_photo'] - 1]
        number_placeholders = location.iloc[i]['position_of_photo']
        #print('i, l, m, номер слайда: ', location.iloc[i]['slide_of_photo'] - 1, "n номер рамки: ", number_placeholders)
        print(location.iloc[i]['position_of_photo'] - 1)
        for k in slide.placeholders:
            print('номер:', k.placeholder_format.idx)
            if number_placeholders == 1 or (location.iloc[i]['slide_of_photo'] - 1 >= 64 and location.iloc[i]['slide_of_photo'] - 1 < 76) or number_ph = patch(slide.placeholders[number_placeholders])
        else:
            ph = patch(slide.placeholders[number_placeholders + 1])
```

## Расширенные отчеты в нужных



# Самые интересные задачи

## Подготовка данных для передачи в HR-report, контроль целостности данных

В HR-report частично отражены показатели Exit-интервью.

Для очистки данных перед их заливкой раньше использовался excel. Я написала скрипт, позволяющий корректно обработать нужные данные в автоматическом режиме. Например, теперь не нужно руками исправлять «Банк Альфа» и «Альфа-банк» среди ответов на вопрос про новое место работы для корректного отображения в HR-report. На текущий момент мною осуществляется коммуникация по поводу корректного отображения данных в HR-report.

С постепенным расширением списка выводимых показателей в HR-report и изменениями скриптов с целью локализации возрастает потребность в корректной обработке данных перед их передачей в HR-report.

## Предиктивная аналитика, построение модели удержания

На текущий момент был перенят опыт построения модели оттока, произведены изменения кода для запуска построенной модели на нашем сервере. Планируется разработка препроцессинга данных.

В связи с запуском процесса удержания возникает потребность в предиктивной аналитике. В текущем спринте был проведен kick-off по модели удержания (задача классификации, успехом считается желание человека уволиться), занимаюсь подготовкой данных (пока кадровых) для создания плоских таблиц, необходимых для корректной работы AutoML.

## СберЧат

После разработки своего чат-бота в рамках курсовой работы в университете хочется внести свой вклад в процесс внедрения чат-бота для целей процесса удержания. Планируется моё непосредственное участие в разработке диаграмм состояний.