

Анализ

*холодных звонков для предсказания
действий клиентов.*

*(покупка автомобильной
страховки)*

Базы данных

Это набор данных из одного банка в Соединенных Штатах.

Помимо обычных услуг, этот банк также предоставляет услуги по страхованию автомобилей. Банк регулярно проводит акции по привлечению новых клиентов. У банка есть данные потенциальных клиентов, и сотрудники банка звонят им для рекламы доступных вариантов автострахования. Нам предоставляется общая информация о клиентах (возраст, работа и т.д.), а также более конкретная информация о текущей кампании продажи страховки (коммуникация, последний день контакта) и предыдущих кампаниях (такие атрибуты, как предыдущие попытки, результат).

Есть данные о 4000 клиентах, с которыми связывались во время последней кампании и для которых известны результаты кампании (покупал ли клиент страховку или нет).

Описание данных

Feature	Description	Example
<i>Id</i>	Уникальный идентификационный номер.	"1" ... "5000"
<i>Age</i>	Возраст клиента	
<i>job</i>	Работа клиента	
<i>Marita</i>	Семейное положение клиента	"разведен", "женат",
"холост"		
<i>Education</i>	Уровень образования клиента	"первичный",
"вторичный"		
<i>Default</i>	Есть ли просроченный кредит ?	да -1, нет -0
<i>balance</i>	Среднегодовой баланс, в долларах США	
<i>HHInsurance</i>	Есть ли страховка?	да -1, нет -0
<i>CarLoan</i>	Имеет ли клиент автокредит	да -1, нет -0
<i>Communication</i>	Тип контактной связи	"сотовая связь",
"телефон",		
<i>LastContactMonth</i>	Месяц последнего контакта	
<i>LastContactDay</i>	День последнего контакта	
<i>CallStart</i>	Время начала последнего вызова (ЧЧ:ММ:СС)	
<i>CallEnd</i>	Время окончания последнего вызова (ЧЧ:ММ:СС)	
<i>NoOfContacts</i>	Количество контактов, выполненных в ходе данной кампании для данного клиента	
<i>DaysPassed</i>	Количество дней, прошедших после того, как с клиентом в последний раз связывались из предыдущей кампании	(числовое значение; -1 означает,
что с клиентом ранее не связывались)		
<i>PrevAttempts</i>	Количество контактов, выполненных до этой кампании и для данного клиента	
<i>Outcome</i>	Результаты предыдущих маркетинговых кампаний	"неудача", "другое",
"успех",		
<i>Carinsurance</i>	Застрахует ли клиент автомобиль?	да -1, нет -0

Что хотим получить?

Задача состоит в том, чтобы построить модель (на выборке 4 000 клиентов), которая бы предсказывала с приемлемой долей вероятности наступит ли нам событие или нет (покупка страховки). В наличии есть данные по 1000 клиентам, по которым сотрудники сделали холодные звонки. Нам нужно на основании данных определить перспективных клиентов и делать уже горячие звонки, с конкретными предложениями.

ДЛЯ КОГО ДЕЛАЕМ?

- Банки
- Страховые группы

Деньги

Имея прогноз покупки страховки клиентом мы можем сконцентрироваться только на тех клиентах, которые по нашей модели готовы купить страховку. Экономия денег компании и времени сотрудников.

Анализ и преобразование данных

Исходные данные:

```
df.dtypes
```

```
Age          int64
Job          object
Marital      object
Education    object
Default      int64
Balance      int64
HHInsurance  int64
CarLoan      int64
Communication object
LastContactDay int64
LastContactMonth object
NoOfContacts int64
DaysPassed   int64
PrevAttempts int64
Outcome      object
CallStart    object
CallEnd      object
CarInsurance int64
dtype: object
```

```
df.isnull().sum().sort_values()
```

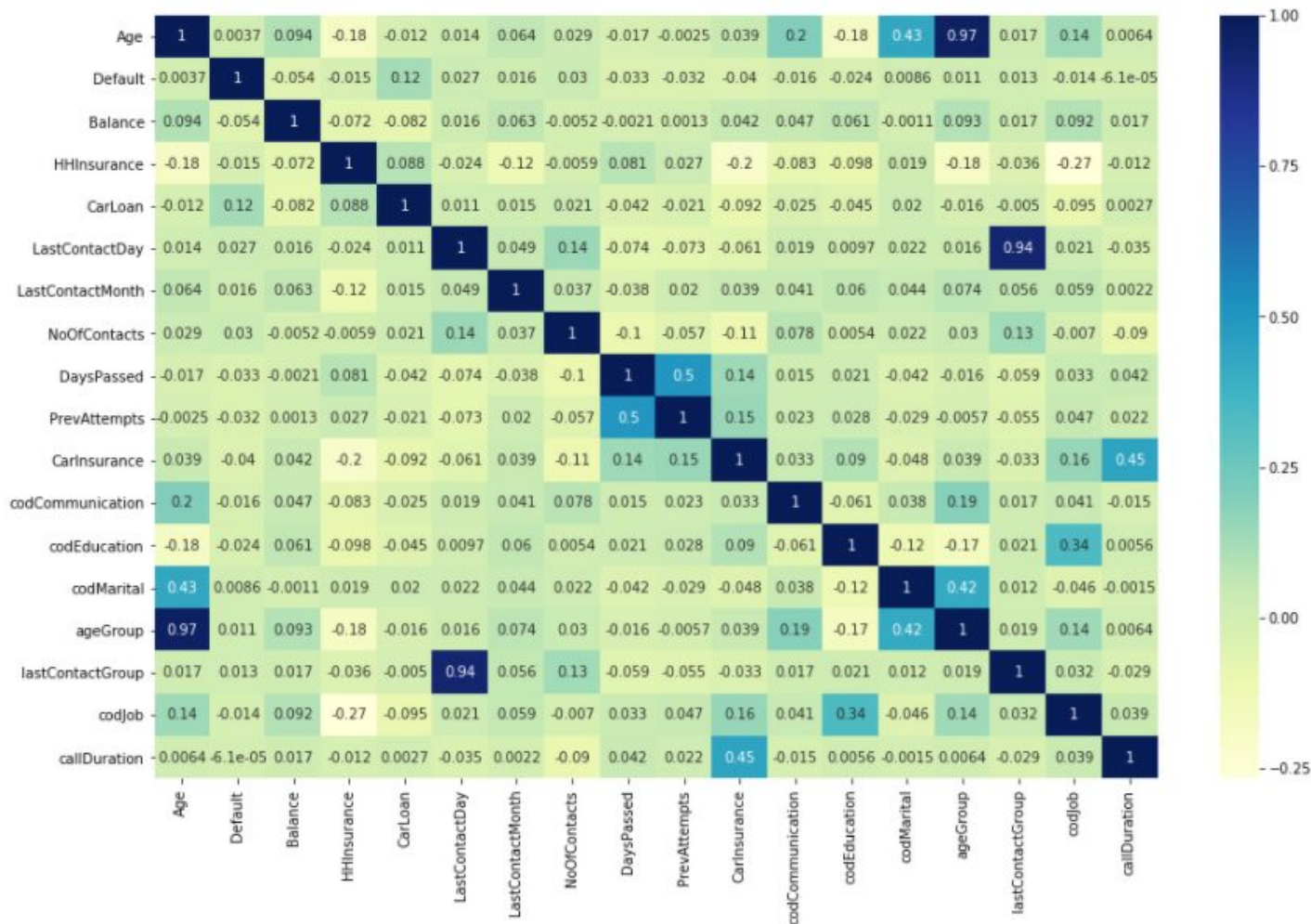
```
Age          0
CallStart    0
PrevAttempts  0
DaysPassed   0
NoOfContacts  0
LastContactMonth 0
LastContactDay 0
CallEnd      0
CarInsurance  0
HHInsurance  0
Balance      0
Default      0
Marital      0
CarLoan      0
Job          19
Education    169
Communication 902
Outcome      3042
dtype: int64
```

Очень много текстовых данных. Будем их преобразовывать.
В столбце Outcome много пропусков- удалим его сразу.

Что меняем...

- Поменяем тип в дате на datetime64
- В 'Job','Education','Communication' вместо неопределенных данных ставим - 0.
- Создаем группы по коммуникации(cellular-1, telephone-2, Nan-0)
- Создаем группы по образованию(primary-1, secondary-2, tertiary -3)
- Создаем группы по семейному положению(primary-1, secondary-2, tertiary -3)
- Создаем группы по возрасту
- Создаем группы по времени от последнего контакта
- Создаем группы по работе клиентов
- Создаем группы по продолжительности общения
- Группы по месяцу последнего общения

График зависимости признаков



Чистим данные

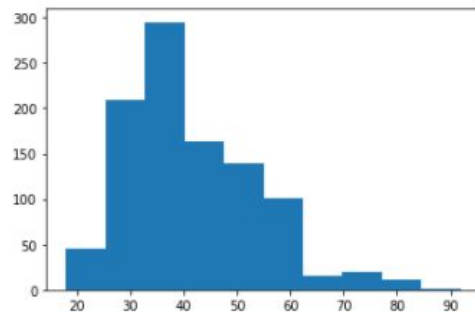
Удалим столбцы с данными, зависимость по которым слабая.

```
df_new=df.drop(columns=['Job', 'Marital',  
'Education',  
'Communication','CallStart','CallEnd'])
```

Распределение данных после обработки (данные распределены равномерно, выбросов нет)

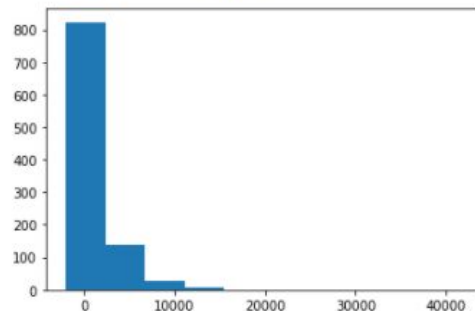
```
plt.hist(test['Age'], bins = 10)
```

```
(array([ 45., 209., 295., 163., 140., 101., 16., 19., 11., 1.]),  
array([18., 25.4, 32.8, 40.2, 47.6, 55., 62.4, 69.8, 77.2, 84.6, 92. ]),  
<a list of 10 Patch objects>)
```



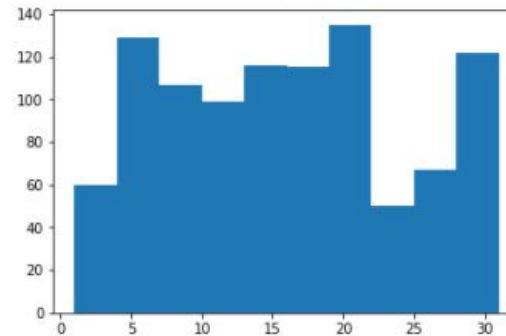
```
plt.hist(test['Balance'], bins = 10)
```

```
(array([823., 139., 29., 6., 0., 1., 1., 0., 0., 1.]),  
array([-1980., 2381., 6742., 11103., 15464., 19825., 24186., 28547.,  
32908., 37269., 41630.]),  
<a list of 10 Patch objects>)
```



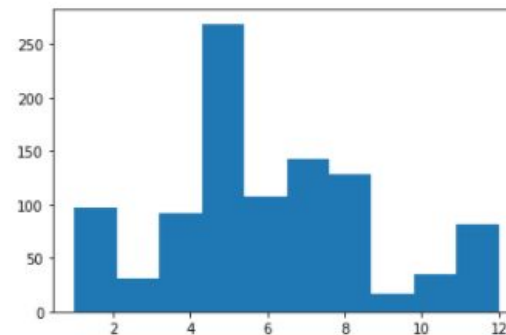
```
plt.hist(test['LastContactDay'], bins = 10)
```

```
(array([ 60., 129., 107., 99., 116., 115., 135., 50., 67., 122.]),  
array([ 1., 4., 7., 10., 13., 16., 19., 22., 25., 28., 31.]),  
<a list of 10 Patch objects>)
```



```
plt.hist(test['LastContactMonth'], bins = 10)
```

```
(array([ 97., 31., 92., 269., 107., 142., 128., 17., 35., 82.]),  
array([ 1., 2.1, 3.2, 4.3, 5.4, 6.5, 7.6, 8.7, 9.8, 10.9, 12. ]),  
<a list of 10 Patch objects>)
```



Построение модели.

Делим данные на Train и Test(25 %)

Используем модель:

GradientBoostingClassifier

test_size = 0.25, random_state = 100

Анализируем полученные данные с исходными: совпадение 80,23%

Приступаем к
прогнозированию на выборке 1
000 клиентов

- Делаем преобразование
данных как при построении
модели
- Применяем выбранную модель
на исследуемых данных с
прогнозом - купит (1) или не
купит страховку(0)

Анализ результатов.

'CarInsurance' = 1 - осталось 527 строк.

Вывод: Анализируя данные 1 000 клиентов банка, по которым были сделаны холодные звонки, приходим к выводу что только 527 клиентов из этого списка купят автомобильную страховку. Данные основаны на построенной нами модели с использованием GradientBoostingClassifier с вероятностью 80,23%

Результат работы

Записываем полученный
результат с прогнозом в файл:

`IVoznenko_car_insurance.csv`