

# Задание банка открытие

## 1 Анализ

Считаем данные и подготовим их к анализу, приведя к чистому виду. (См. приложение "Подготовка данных")

В таблице нет информации за первые месяцы от банков **D** и **E** - возможно, они недавно открылись. Более примечательно, что от банка **E** нет информации за последние месяцы. Возникает предположение, что это отделение закрылось. Посмотрим, есть ли какие-то очевидные поводы для такого решения.

Построим графики абсолютных показателей (количество поданных заявок, одобренных заявок и выданных кредитов):

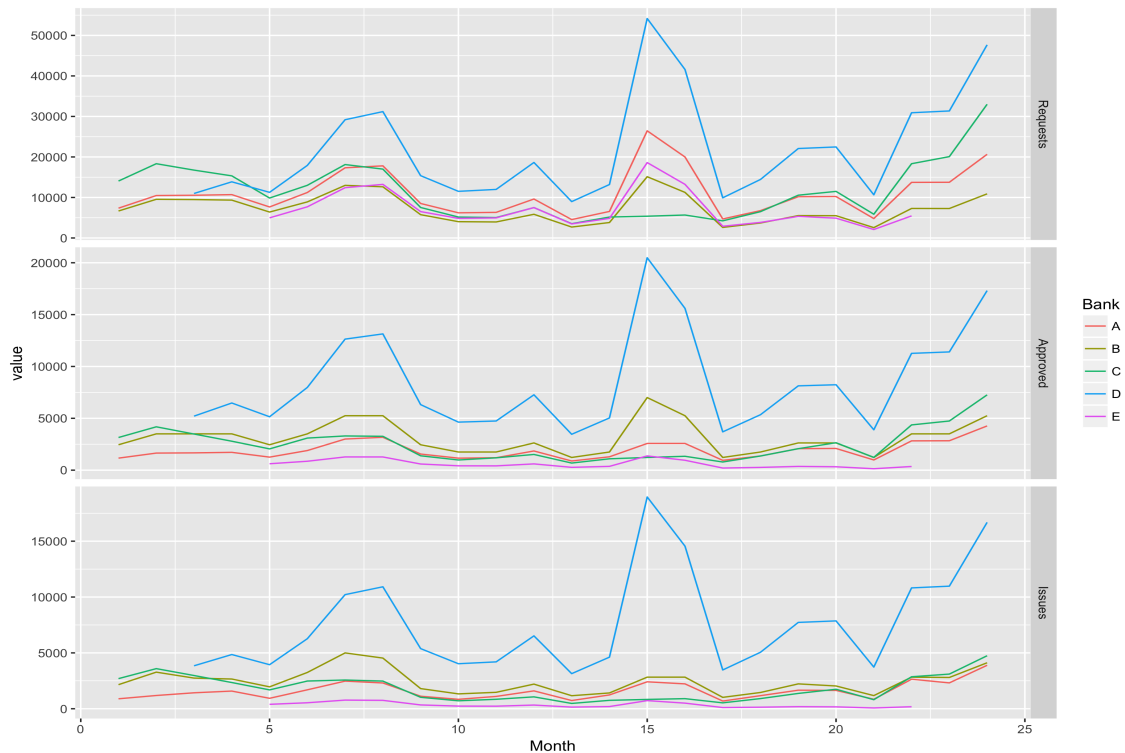


Рис. 1: Графики абсолютных показателей.

Действительно, в банк **E** долгое время подавалось наименьшее количество заявок, почти всегда там одобрялось наименьшее количество заявок и всегда там выпускалось наименьшее количество кредитов. Возможно, он находился в районе с малым числом потенциальных клиентов.

Чтобы не нарушать предпосылок таких тестов, как ANOVA и тест Тьюки (см. приложение "Распределение абсолютных показателей"), воспользуемся непараметрическим тестом Данна, чтобы узнать, в какие банки подавалось статистически больше заявок. (см. приложение "Результаты статистических тестов") Наибольшее отличие от всех остальных видно в банке **D**. Менее всего различаются пары **A** и **C**, а так же **B** и **E**. Возможно, другие банки проводили более гибкую кредитную политику. Посмотрим на общее количество абсолютных показателей за всё время:

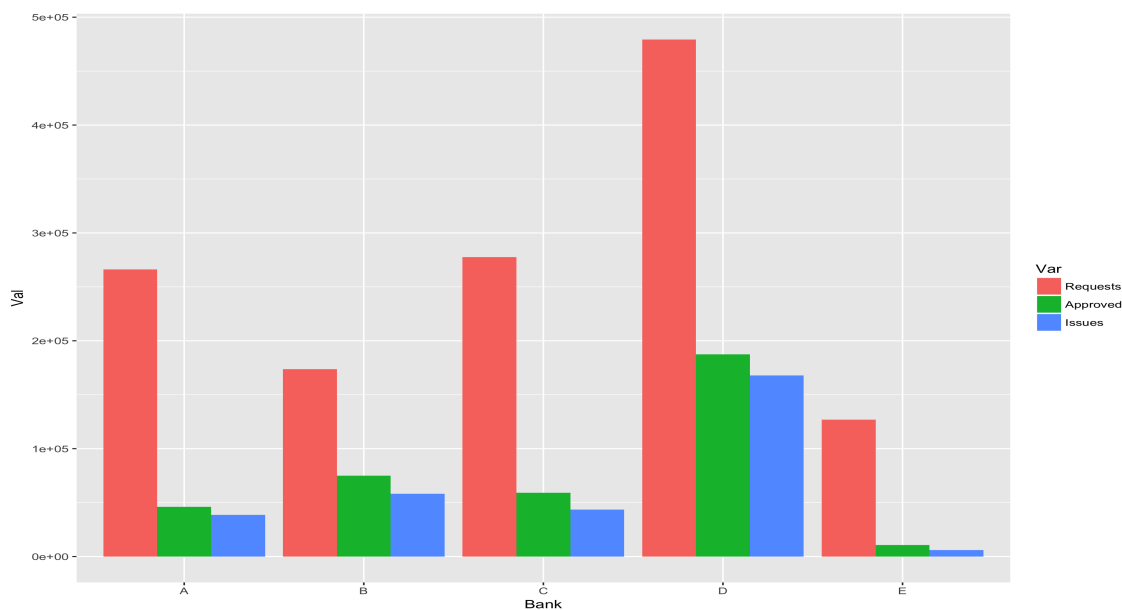


Рис. 2: Сумма абсолютных значений за 24 месяца.

Действительно, в банк **Е** было подано на 27% заявок меньше, чем в банк **В**, а в банки **А** и **С** соответственно на 53% и 60% больше, однако, число одобренных и выданных кредитов в банке **В** даже больше, чем в последних. Взглянем на относительные уровни одобрения и согласия (проценты одобренных заявок из общего количества и выданных кредитов из количества одобренных соответственно).

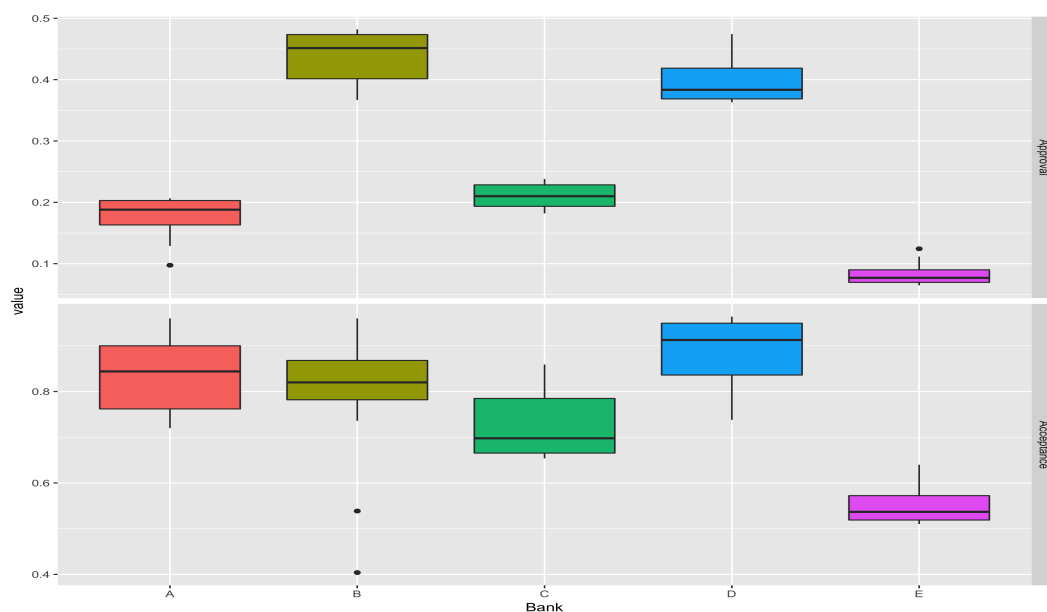


Рис. 3: Относительные уровни одобрения и согласия.

В самом деле, банк **В** одобрял в среднем даже больше самого успешного банка **Д**. Возможно, последний находится в густонаселенном районе и подобная политика привела к лояльности клиентов к банку, что и стало причиной его успеха. Тогда банку **В** стоит продолжать придерживаться данной стратегии, если это не приводит к росту невыплаченных задолженностей. Продолжим анализ и посмотрим, какие из этих различий статистически значимые (см. приложение "Результаты статистических тестов").

По уровню одобрения, банки **В** и **Д** действительно похожи, а предположительно закрытый банк **Е** сильно отличается от всех остальных. Больше всего с ним статистически схож банк **А**. По уровню согласия на одобренные условия банк **Е** уступает всем остальным. Однако и в банке **С** клиенты берут одобренные кредиты статистически меньше, чем в банках **А**, **В** и **Д**. Эти замечания являются поводом для пересмотра

кредитной политики в банках **А** и **С**. Такая строгость при одобрении может послужить причиной потери лояльности клиентов и повторения судьбы банка **Е**.

## 2 Резюме

Исходя из приведённого анализа можно сделать следующие выводы:

- Если гибкая кредитная политика банков **В** и **Д**, характеризующаяся высоким процентом одобрения заявок, не приводит к росту задолженности клиентов и прибыльности банка, но вызывает приток лояльных клиентов, её можно считать успешной.
- Строгая кредитная политика банка **Е** приводит к оттоку клиентов и малой популярности. В связи с этим, банкам **А** и **С** следует обратить внимание на свои порядки одобрения заявок.
- Анализ следовало бы продолжить, опираясь на дополнительные данные. Определение эффективности и выбор модели одобрения должен быть результатом анализа таких показателей, как параметры региона, макроэкономической ситуации и даже самого банка.
- Так же в дальнейшем анализе потребуются пропущенная информация, возможно, вычисленная по некоторой модели (если её нельзя добыть, например, потому что эти банки в эти месяцы были закрыты), чтобы мы имели полную картину хотя бы за два полных периода. Тогда будет полезно сделать попарные сравнения первых и последних 12 месяцев для каждого банка, чтобы сделать вывод о последствиях их кредитной политики. Возможно, негативные эффекты уже сказываются на банках **А** и **С**.
- Более того, имея информацию как минимум о двух периодах, было бы полезно проанализировать данные, как временные ряды. Возможно, исходя из анализа сезонности, мы смогли бы объяснить отличия в видах линий, пики и их разнонаправленность, которые мы видим на графиках показателей, нормированных на отрезок  $[0,1]$ .

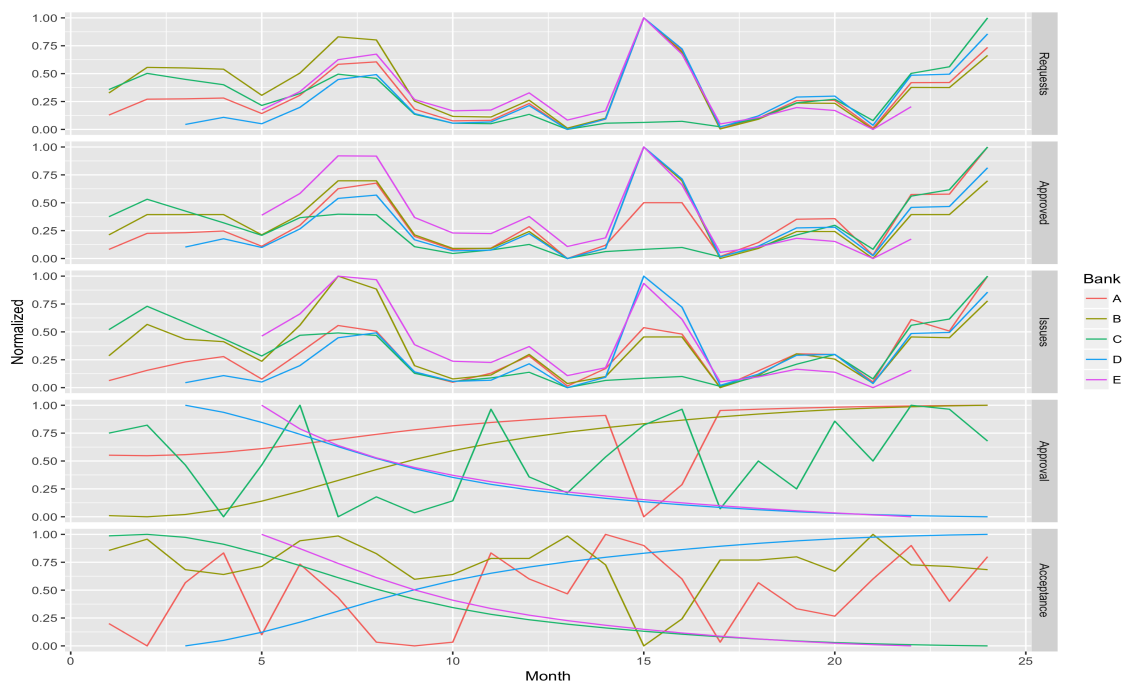


Рис. 4: Нормированные показатели

### 3 Приложения

#### 3.1 Подготовка данных

Для начала считаем данные из файла и взглянем на них.

Bank Parameter		M1	M2	M3	M4
A	Кол-во принятых кредитных заявок	7357.1012019	1.047682e+04	1.054394e+04	1.070278e+04
NA	Уровень одобрения	0.1576522	1.571524e-01	1.581591e-01	1.605418e-01
NA	Кол-во выданных кредитов	890.7747526	1.185449e+03	1.427483e+03	1.580785e+03
B	Кол-во принятых кредитных заявок	6660.2318842	9.544878e+03	9.484121e+03	9.343363e+03
NA	Уровень одобрения	0.3678551	3.666888e-01	3.690379e-01	3.745975e-01
NA	Кол-во выданных кредитов	2156.0000000	3.276000e+03	2.744000e+03	2.660000e+03

Приведём данные к *чистому* виду. Каждое наблюдение будет в своём ряду, каждый показатель - в своём столбце.

Bank Month Requests Approval Issues				
A	1	7357	0.158	891
A	2	10477	0.157	1185
A	3	10544	0.158	1427
A	4	10703	0.161	1581
A	5	7657	0.164	935
A	6	11229	0.168	1695

Данные отсортированы по месяцам внутри каждого отдельного банка. В конце таблицы будут данные за последние месяцы пятого банка, который мы обозначили за **Е**.

### **Bank Month Requests Approval Issues**

E	19	5325	0.068	186
E	20	4891	0.067	168
E	21	2090	0.066	70
E	22	5455	0.065	180
E	23	NA	NA	NA
E	24	NA	NA	NA

Тут же мы видим, что некоторые данные в таблице отсутствуют. Всего строк с пропущенными значениями в файле 8. Это не так много, поэтому взглянем на них.

### **Bank Month Requests Approval Issues**

D	1	NA	NA	NA
D	2	NA	NA	NA
E	1	NA	NA	NA
E	2	NA	NA	NA
E	3	NA	NA	NA
E	4	NA	NA	NA
E	23	NA	NA	NA
E	24	NA	NA	NA

Для эффективности анализа добавим в нашу таблицу столбцы с абсолютным количеством одобренных заявок, а так же "уровень согласия то есть процент одобренных заявок, на которые согласились клиенты.

### **Bank Month Requests Approved Issues Approval Acceptance**

A	1	7357	1160	891	0.158	0.768
A	2	10477	1646	1185	0.157	0.720
A	3	10544	1668	1427	0.158	0.856
A	4	10703	1718	1581	0.161	0.920
A	5	7657	1256	935	0.164	0.744
A	6	11229	1891	1695	0.168	0.896

### 3.2 Распределение абсолютных показателей

Так выглядят гистограммы наших абсолютных показателей:

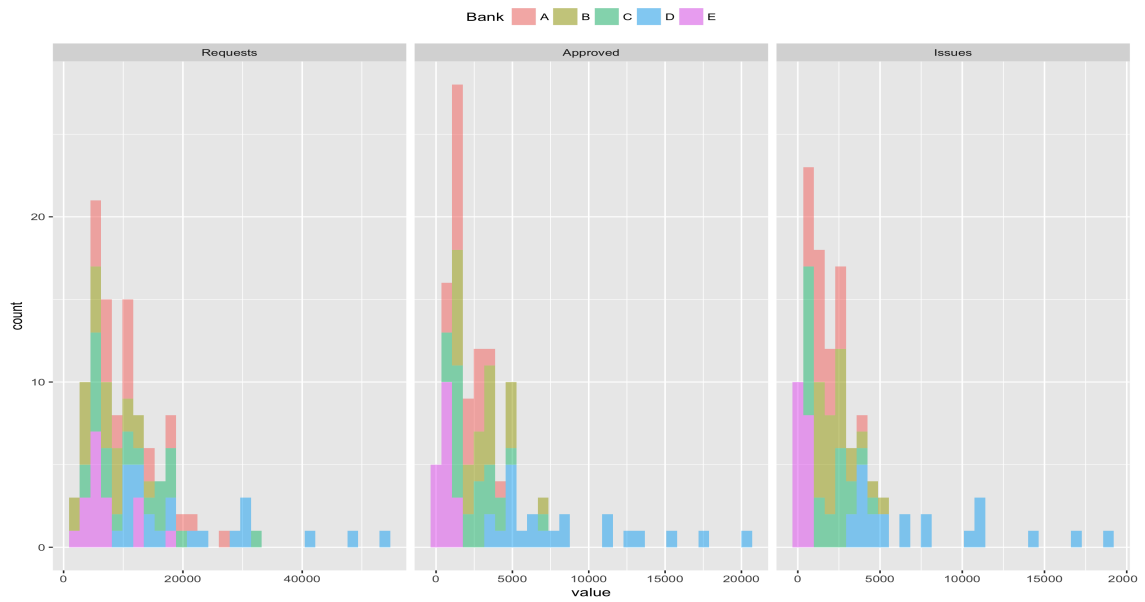


Рис. 5: Гистограммы абсолютных показателей.

Видно, что они имеют далеко не нормальное распределение. Простое логарифмирование тоже нормализует слишком грубо:

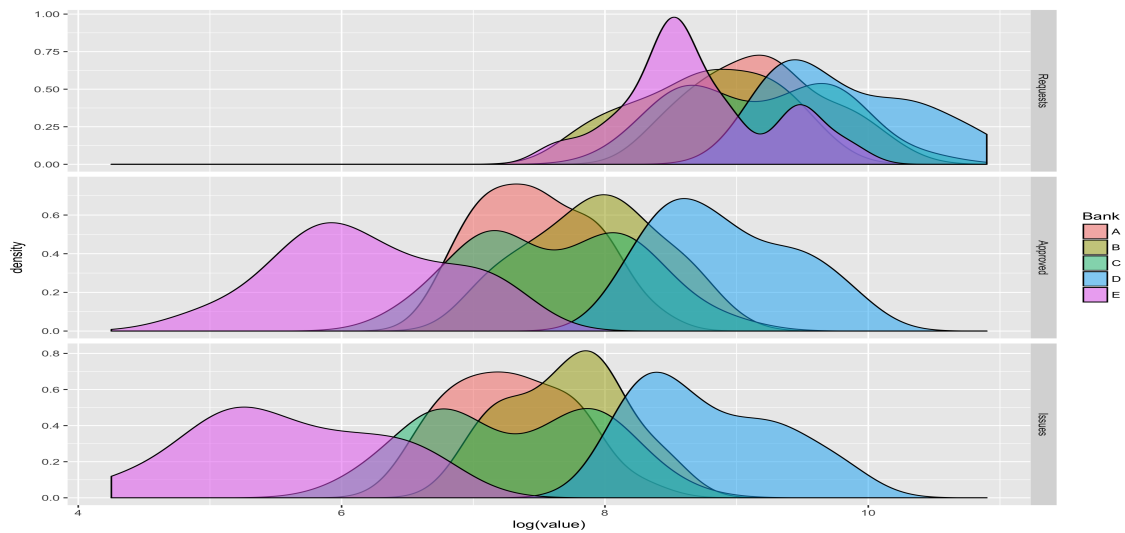


Рис. 6: Плотности логарифмов.

Хотя есть ряд работ, утверждающих, что дисперсионный анализ достаточно устойчив к нарушению условия нормальности распределения, лучше воспользуемся непараметрическим тестом Данна.

### 3.3 Результаты статистических тестов

Kruskal-Wallis rank sum test

data: x and group

Kruskal-Wallis chi-squared = 37.0577, df = 4, p-value = 0

Comparison of x by group (Bonferroni)					
Col Mean-I					
Row Mean I		A	B	C	D
B		2.231156			
		0.1284			
C		-0.011111	-2.242268		
		1.0000	0.1247		
D		-3.063733	-5.245847	-3.052866	
		0.0109	0.0000	0.0113	
E		2.366033	0.300383	2.376320	5.166562
		0.0899	1.0000	0.0874	0.0000

Рис. 7: Результаты теста на общее количество заявок.

Kruskal-Wallis rank sum test

data: x and group

Kruskal-Wallis chi-squared = 98.3324, df = 4, p-value = 0

Comparison of x by group (Bonferroni)					
Col Mean-I					
Row Mean I		A	B	C	D
B		-6.475743			
		0.0000			
C		-1.666715	4.809028		
		0.4779	0.0000		
D		-5.060952	1.272450	-3.430872	
		0.0000	1.0000	0.0030	
E		2.464810	8.460184	4.007888	7.118304
		0.0685	0.0000	0.0003	0.0000

Рис. 8: Результаты теста на уровень одобрения.

Kruskal-Wallis rank sum test

data: x and group

Kruskal-Wallis chi-squared = 64.0496, df = 4, p-value = 0

		Comparison of x by group (Bonferroni)			
Col	Mean-I				
Row	Mean	A	B	C	D
-----+-----					
B		0.404470			
		1.0000			
C		2.997973	2.593502		
		0.0136	0.0475		
D		-1.672621	-2.068201	-4.604697	
		0.4720	0.1931	0.0000	
E		5.787782	5.413315	3.012198	7.231602
		0.0000	0.0000	0.0130	0.0000

Рис. 9: Результаты теста на уровень согласия.