

Какая игра будет успешна на кикстартере?

Гипотетический заказчик:

Группа энтузиастов из США с идеей для игры, решающая в каком формате им стоит реализовать свою задумку.

Ожидания заказчика:

Получить данные, которые потенциально смогут помочь ему запустить успешную компанию по сбору денег на игру. Понять тенденции в сборе денег для игр, понимание того, на что стоит обращать внимание при запуске.

Способ реализации:

Изучение датасета с данными о проектах с кикстартера. Путем изучения данных обнаружить статистически значимые переменные в условиях успешности проекта. Провести статистические тесты чтобы удостовериться в точности предсказаний.

Исследовательские вопросы:

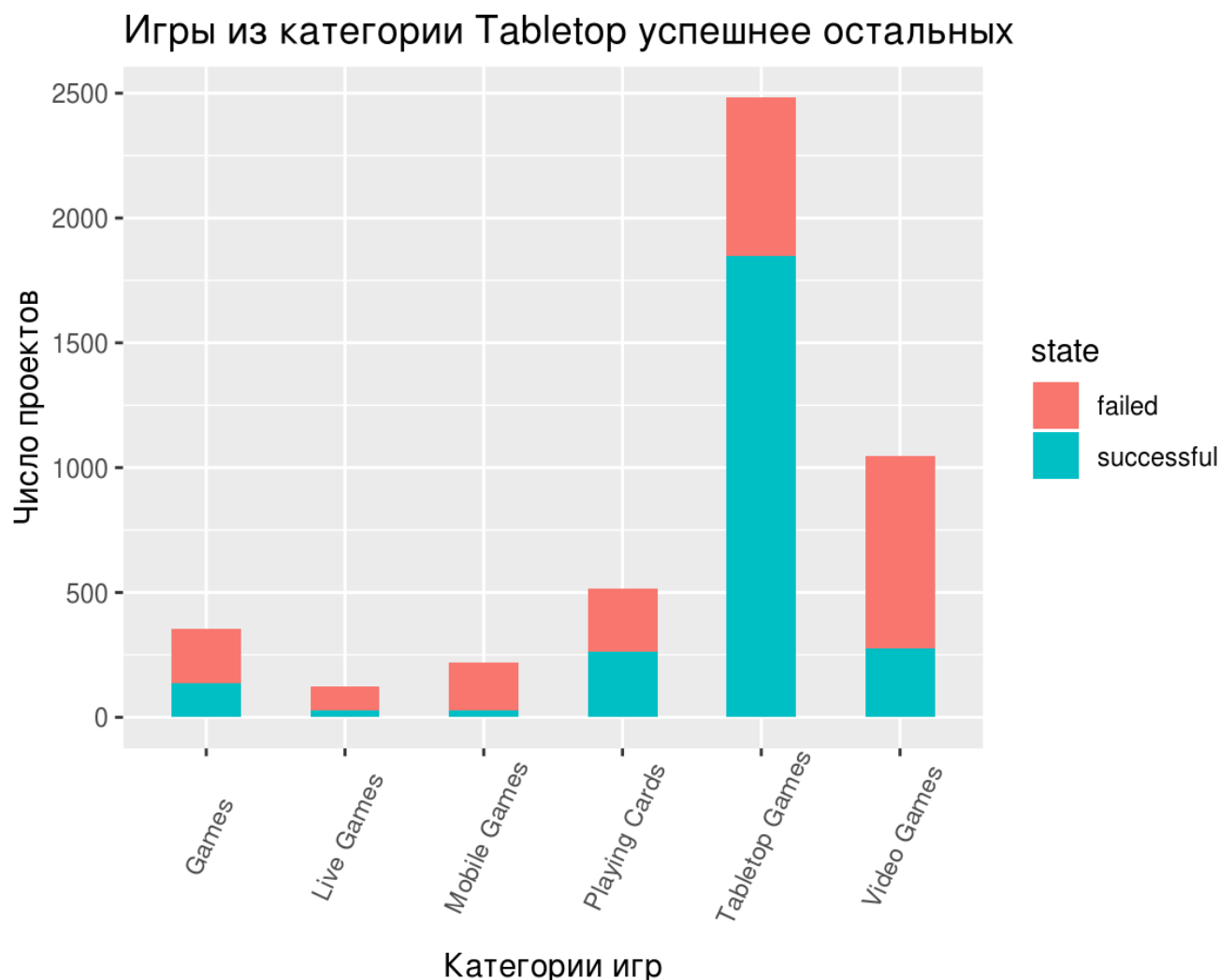
- Влияет ли страна запуска проекта в категории игр на успешность?
- Какой вид игр наиболее успешен?
- Существует ли зависимость в требуемой сумме между успешными и неуспешными проектами в категории игр?

Влияет ли страна на успешность?



Как можно заметить, количество проектов, в том числе успешных, гораздо больше в США, чем в других странах. Поэтому заказчик, будучи из США - может переходить к следующей стадии - решению с помощью какого вида игры реализовать проект.

Какой вид игр наиболее успешен?



Можно заметить, что игры категории Tabletop значительно популярнее и успешнее остальных. На всякий случай, проверим результат с помощью хи-квадрата.

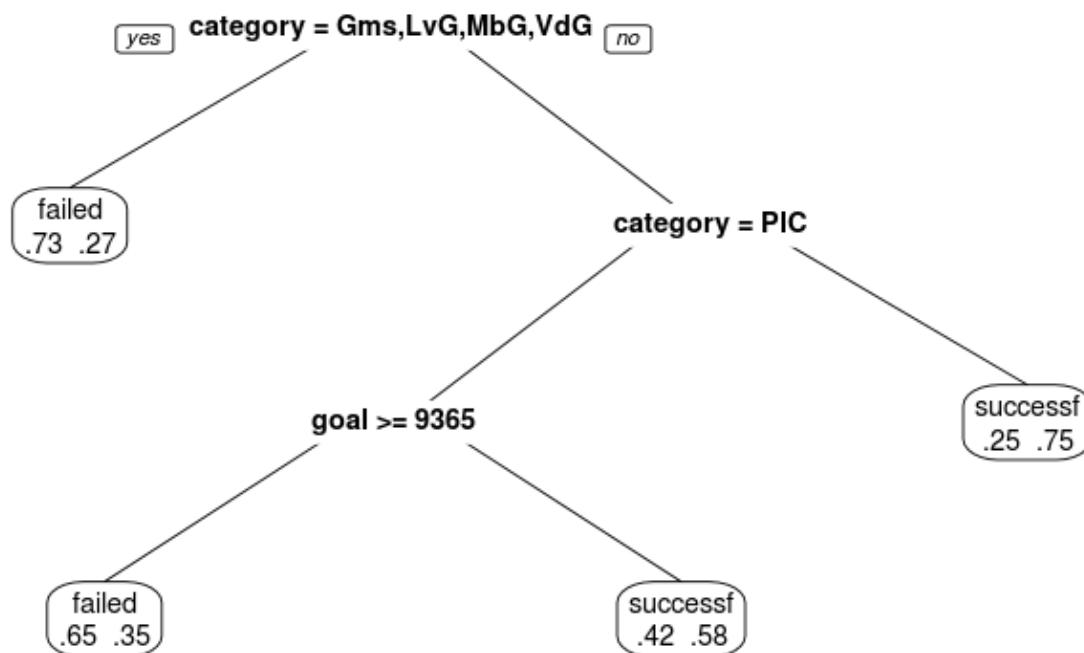
- Есть ли связь между категорией игры и успешностью?
- H_0 : Между категорией игры и успешностью игры нет взаимосвязи.
- H_1 : Между категорией игры и успешностью взаимосвязь есть.

Pearson's Chi-squared test

```
data: kickstarter_games2$category and kickstarter_games2$state  
X-squared = 971.74, df = 5, p-value < 0.00000000000000022
```

- p-value крайне низок, соответственно - взаимосвязь есть.

На случай, если заказчик не захочет остаться с единственным вариантом - проведем тест с деревом решений.



Можно пронаблюдать, что если откинуть самый выигрышный вариант с настольными играми, наиболее привлекательным останется вид “Playing Cards”. Проверим точность построенного древа.

```

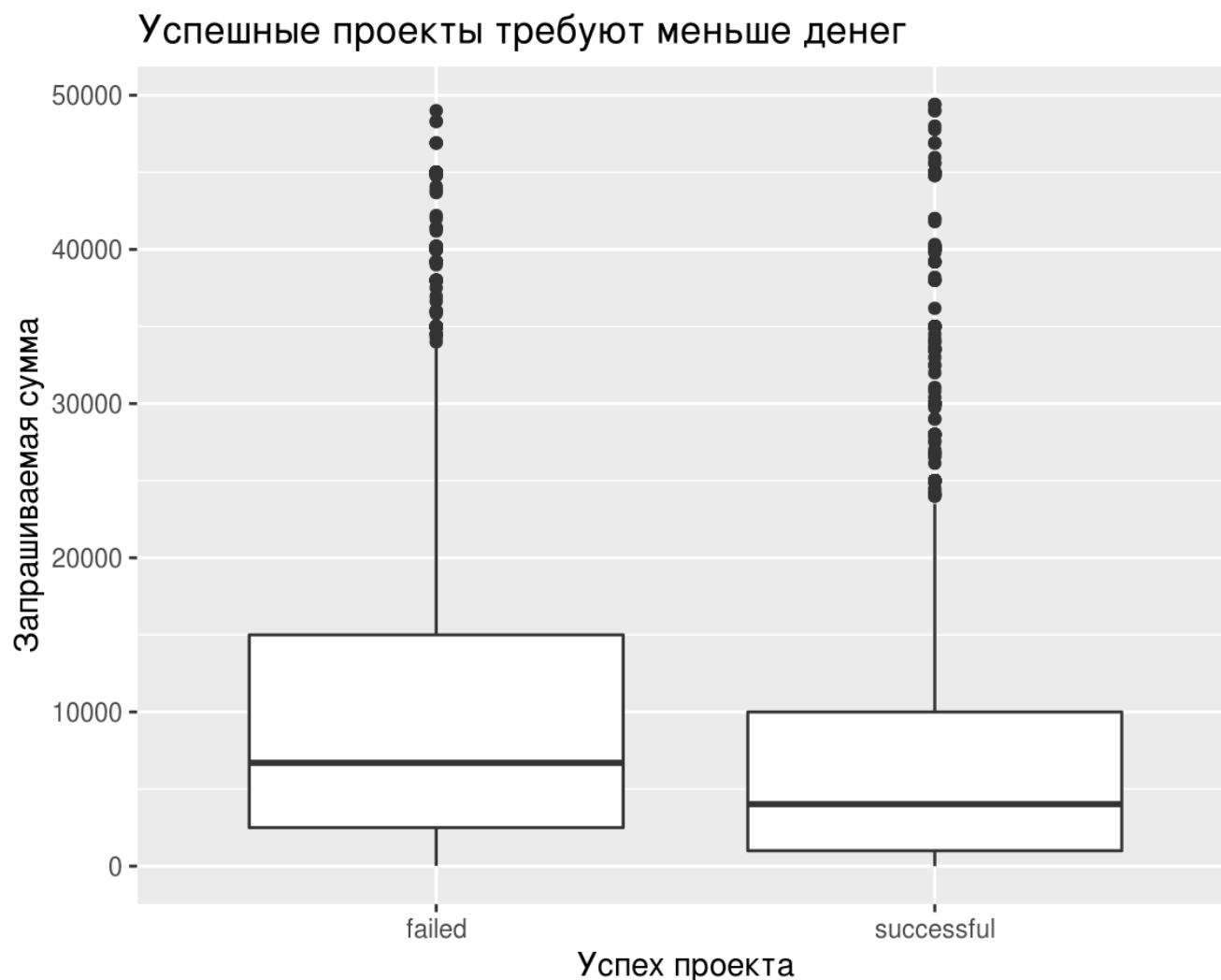
pred = predict(tree.kick, type="class")
t = table(pred, kickstarter_games2$state)
t
(t[1,1] + t[2,2])/sum(t)
~~~~

```

```
[1] 0.723687
```

Построенная модель имеет точность в 72%, что достаточно для того, чтобы посудить, что результат не случаен и категория карточных игр вторая по привлекательности для проекта.

Существует ли зависимость в требуемой сумме между успешными и неуспешными проектами в категории игр?



Определившись с типом игры, нужно узнать какой следует установить бюджет для успешного запуска. Из боксплота можно увидеть, что успешные проекты обычно требуют меньше денег. В среднем это 8689\$. Чтобы удостовериться, можно провести т-тест.

- H_0 : Статистической разницы в запрашиваемой сумме между успешными и неуспешными проектами нет.
- H_1 : Статистической разницы в запрашиваемой сумме между успешными и неуспешными проектами есть.

Welch Two Sample t-test

```
data: goal by state
t = 10.236, df = 4367.6, p-value < 0.00000000000000022
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 2331.075 3435.516
sample estimates:
 mean in group failed mean in group successful
      10257.444          7374.148
```

С помощью т-теста мы выяснили, что p-value крайне мал, поэтому можно отвергнуть H_0 , а значит статистическая разница в запрашиваемой сумме между успешными и неуспешными проектами есть - наше предположение верно.

Результат

- Существует взаимосвязь между страной игрового проекта и успешностью сбора денег. Т.к заказчик из США, их игре будет легче реализоваться.
- Существует зависимость успешности игры и ее типом. Наиболее успешны и привлекательны для заказчика - настольные и карточные игры.
- Сумма, запрашиваемая для донатов влияет на успешность. Следовательно, запрашиваемую сумму для проекта стоит выбирать в районе средней для успешных - 8689\$.