# Проєкт

У цьому файлі стисло передано сутність того, що я робив у цьому проєкті, код якого можна отримати за посиланням [[1]](#сил1) (щоб знайти зазначену інформацію тут требі інколи погортати вниз у зазначеній частині коду), а візуалізацію детальну за [[2]](#табл).

## Завдання

На основі даного датасету для аналізу європейського регіону зробити наступні дії:

1. Сегментувати клієнтів;

2. Порахувати LTV (врахувати, що 25% замовлень повертають, що в грошовому еквіваленті складає 8.4% від усієї суми);

3. Дослідити активність найкращих клієнтів і запропонувати товари для Basketing Analysis;

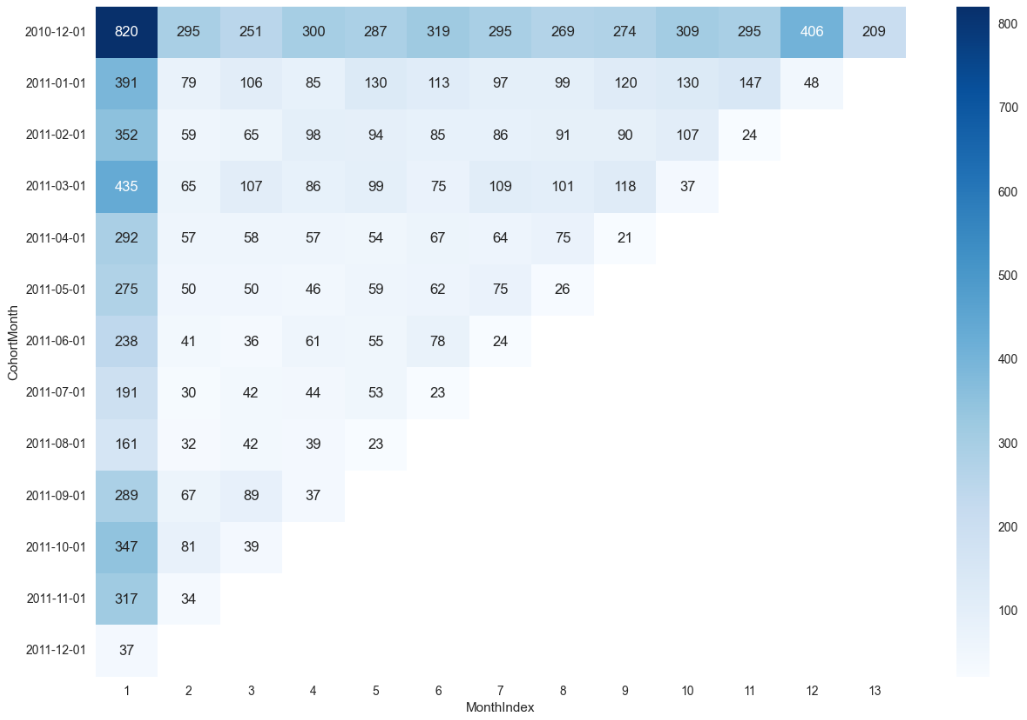
4. Все, що було досліджено, зобразати в Tableau.

## Когортний аналіз

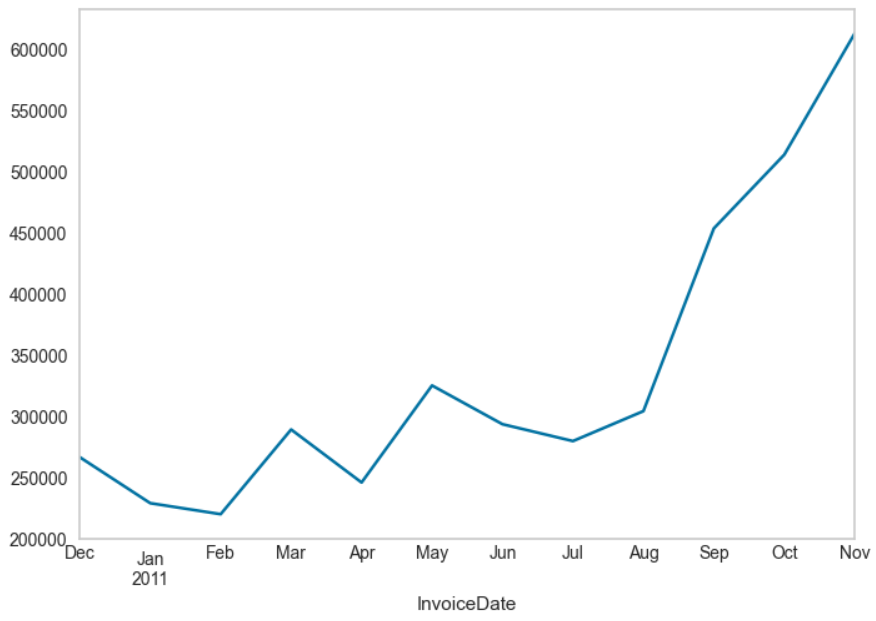
У цій частин блоку я вивчав на активність нових юзерів щомісячно [Рис.1.](#рис11) У перший місяць найбільша кількість людей приєдналась. Також спостерігаємо, що найменша активність літом, яка «зловила» цей тренд ще у квітні й відновилася осінню. Середня кількість користувачів, які приєднуються кожного місяця, складає 277.

Оскільки кінець весни та літо (з економічної точки зору) не є активними, та й нові користувачі в цей період активно не приєднуються – масштабні кампанії немає сесну проводити. Натомість можна созередити увагу на лояльних користувачах та локальних кампаніях для новеньких.

На [Рис. 2.](#рис22) бачимо, що літом теж відбувається невеликий спад в продажах, але загальний тренд на збільшення наявний, що й бачимо з настанням вересня.



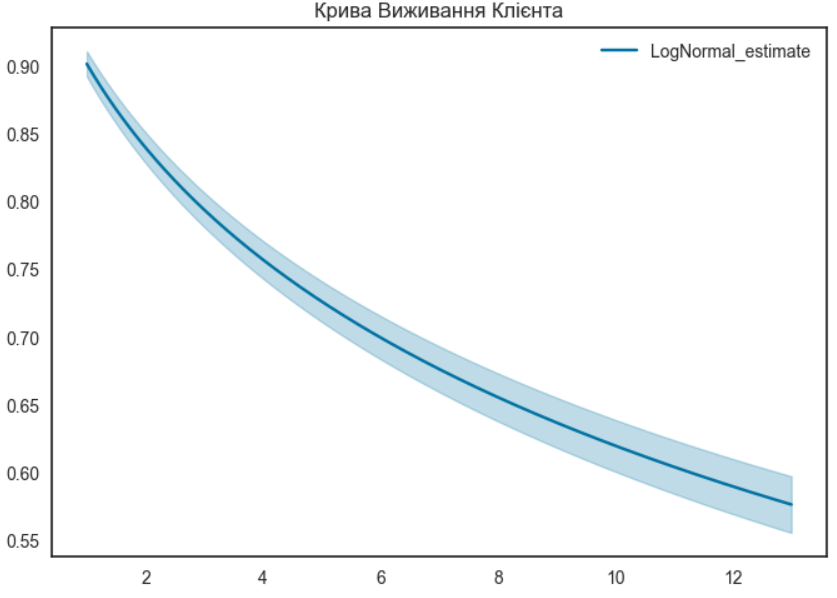
1. Щомісячні когорти.



1. Динаміка продажів по місяцям.

## Скільки живе клієнт та Churn Rate

На основі Survival Analysis я знайшов, що в середньому наш клієнт живе приблизно 20 місяців [[1, “Survival analysis”]](#сил1). А «Churn Rate» скаладє **33.43%**.



1. Крива виживання.

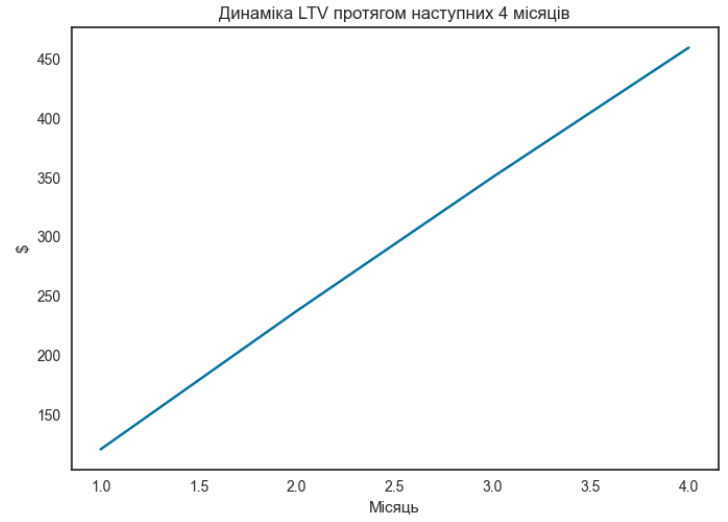
## LTV

Для прогнозу LTV я використовував декілька методик розрахунку (на забуваючи про те, що повертають):

* *агрегований* (3394.6$);
* за роботою *Survival Curve* (життя клієнта ~18 місяців, за розрахунком складає 3353.6$);
* *ймовірнісна модель* (461.31$ на наступні 4 місяці (без урахування теперішної цінності клієнта), похибка складає 366.00$) на [Рис. 4.](#рис2);
* та *регресію*, для якої я моделював на наступні 3 місяці (1084.42$, у якої похибка = 214.3$). [[1, “LTV :)” та “ML :)”]](#сил1)

З усіх наявних результатів я б для короткострокової перспективи використовував результати роботи регресії, бо модель ця найточніша та й не викривлена результатами високих сегментів, де LTV клієнта може бути занчено більше 5000$.

Для довгострокової можна використовувати *Survival* та *агрегований*, адже оцінки майже однакові. Щоб клієнти жили з нами більше – треба стимулювати купувати їх більше, адже з кожною новою покупкою їх життя з нами продовжується 1.86 разів.

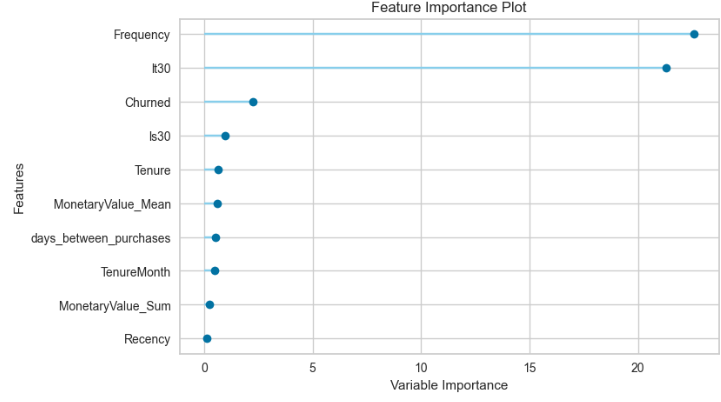


1. Динаміка LTV на 4 місяці за BG/NBD (ймовірнісною) моделлю.

### **Аналіз фаткорів впливу**

За регресійною моделлю [Рис. 5.](#рис3) найважливішими факторами впливу стали:

1. періодичність покупок;
2. скільки клієнт зробив транзакцій у свій останній місяць активності;
3. чи “Churn” клієнт;
4. скільки витратив в останньому місяці;
5. скільки він з нами живе.



1. Фактори, які впливають на майбутні витрати клієнта.

Тобто найголовнішим фактором є періодичність придбань, тому клієнту треба пропонувати найшпоширеніші товари та суміжні для них за зниженими цінами. Для цього я провів Basket Analysis в Python та Tableau [[1, “Basket Analysis ”](#сил1); [2, “Growth Opportunity”](#табл)], де в першому варіанті це зроблено точніше, але не зовсім гарно, а в другому можна обирати товари, яки корелюють між собою й обирати ті продукти для проведення кампаній, що найкраще продаються, щоб в кінцевому результаті не втрачати на масштабі та провокувати клієнтів робити більше придбань.

Також в Tableau [[2, “Sales Exploration”]](#табл), можна дивитися, які товари популярні в конкретну годину та день, щоб теж пропонувати за зниженою ціною товари знаючи, що їх загальні продажі користуються попитом у цей час.

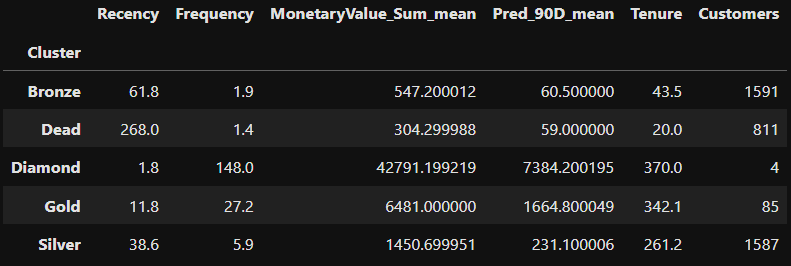
Якщо клієнта постійно зацікавлювати – менша ймовірність його churn.

LTV для кожної країни (який розраховувався як теперішня вартість + прогнозована вартість на наступні 90 днів) можна побачити в Tableau [[2, “Country and Segment Analysis”]](#табл) та як це узагальнене значення змінюється при зміні сегменту, країни.

## Сегментація

Провівши сегментацію, я отримав 5 кластерів: Dead (неактивні клієнти), Bronze (слабкі фінансово клієнти, але не об’ємом грошей з цієї групи), Silver (середній клас), Gold (вище середнього класу) та Suprime (4 особи серед усіх, які приносять дуже багато грошей). Розподіл сегментів у інтерактивному 3D просторі можна побачити тут [[1, “Clustering”]](#сил1) або [тут](#ксла). Також за посиланням [[2, “Country and Segment Analysis”]](#табл) можна вивчити деталі життя кожного кластеру.

Таблично загальні характеристики кластерів виглядають наступним чином:



1. Характеристики кластерів.

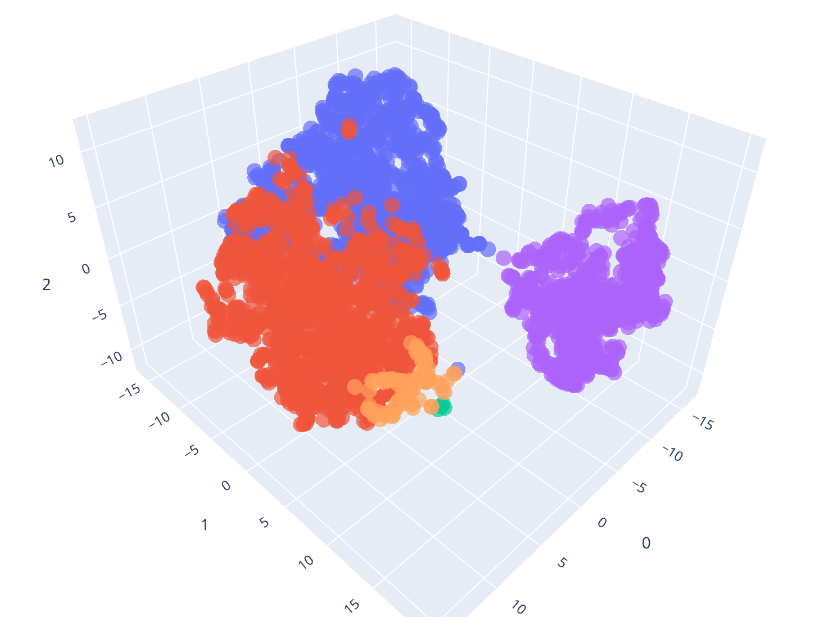
Від 'Bronze' багато неможливо отримати в натупні місяці, але гроші, які зароблені з них за останні місяці, значні в порівнянні з іншими кластерами [[2, “Country and Segment Analysis”]](#табл), тому для них треба проводити окремі кампанії для їх утримування та стимуляції.

Те саме можна сказати за “Silver”, але вони більш активні та більше приносять грошей на клієнта.

Увагу треба зосередити на "Diamond", 'Gold', адже вони приносять значні гроші на одного клієнта.

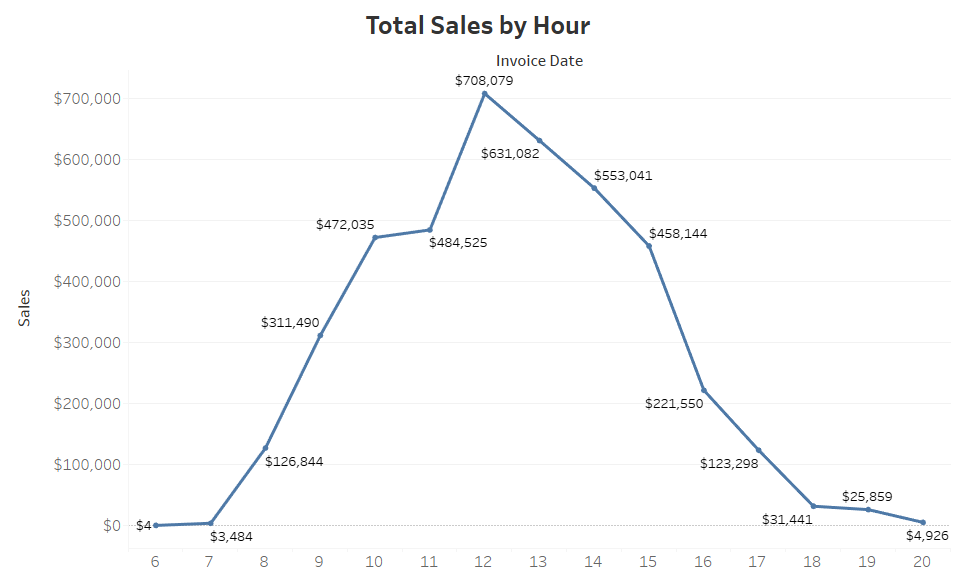
'Dead' - churn клієнти або "плохіши", тому забуваємо за них й не витрачаємо гроші на їх повернення, а краще інвестувати в експансію, за яку мова піде в наступному пункті.

Розподіл кластерів у 3D просторі:



## Оптимізація

По-перше, треба зменшити часи роботи, бо як можна побачити на [Рис. 7.](#рис7), о шостій, сьомій ранку та о восьмій вечора ми за весь час заробили загалом 8414$, що складає ~0.2% від нашого доходу, коли ми для їх реалізації витрачаємо на 3 год більше роботи команди кожного дня.

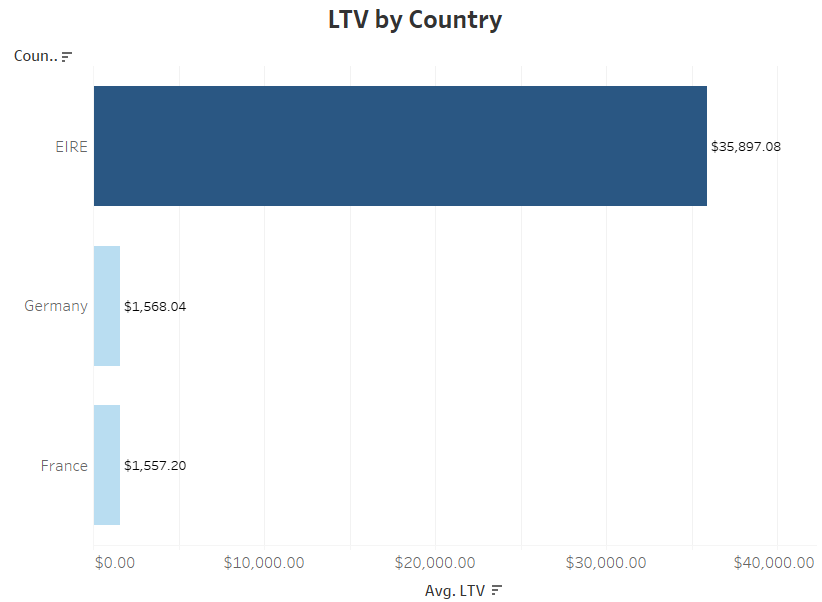


1. Загальна активність клієнтів по годинам.

По-друге, треба проводити експансію, оскільки одна Англія займає ~88% ринку в грошовому еквіваленті й наявні декілька перспективних регіонів, частки ринку яких можна побачити на малюнку:



1. Відсоткове відношення всіх наших грошей серед цікавих для бізнесу регіонів.



1. LTV на даний момент серед покупців привабливих регіонів.

Треба проводити впевнену експансію на наступні 3 регіони, бо їх потенціал видно на малюнках вище (враховуємо, що в Ірландії всього 3 юзера):

* Німеччина;
* Франція;
* Ірландія.

## Висновок

Отже:

1. Треба стимулювати юзерів купувати частіше, адже це підвищує їх довіру до нас. За допомогою “Basket Analysis” розуміти, що треба продавати клієнтам більше;
2. треба провести експансію на території Німеччини (частка ринку 3.68%) та Франції (частка ринку 3.25%), Ірландії (частка ринку 0.46%);
3. зменшити години роботи команди, бо 8414$ за рік за додаткові 3 години кожень день того не варте;
4. проводити масштабні маркетингові кампанії з залученням нових клієнтів та для утримання лояльних (але не влітку).

Для більш детального вивчення цих пунктів рекомендую побачити в Tableau дашборди за посиланням [[2]](#табл).

Дякую за увагу!

Посилання:

1. <https://github.com/Piip218/ccc/blob/segmentation/code.ipynb>; python

1. <https://public.tableau.com/app/profile/vova.doms/viz/Project_17200908940370/FinalStory>. tableau