**Анализ эффективности торгового предприятия Ozon**

Тищенко И.С.

ПМ23-1

**Аннотация:** Данная работа посвящена анализу эффективности маркетплейса Ozon на основе финансовых и операционных показателей. Построена регрессионная модель, позволяющая оценить влияние ключевых метрик на количество покупок за квартал. Проведён анализ адекватности модели, включая проверку гипотез и экономическую интерпретацию коэффициентов. В результате выполнено прогнозирование, рассчитаны доверительные интервалы. Работа завершена выводами о текущей эффективности маркетплейса и рекомендациями для улучшения его работы.

**Ключевые слова:** эффективность, маркетплейс, регрессионная модель, прогнозирование, финансовые показатели, Ozon.

### **Постановка задачи**

Целью работы является оценка эффективности маркетплейса Ozon на основе финансовых и операционных показателей, а также разработка модели прогнозирования количества покупок за квартал.

Для достижения этой цели необходимо:

* Изучить доступные данные, включая финансовые отчёты и операционные показатели.
* Построить регрессионную модель, отражающую взаимосвязи между метриками.
* Провести анализ адекватности модели, включая проверку гипотез и экономическую интерпретацию коэффициентов.
* Выполнить прогнозирование основных показателей и оценить доверительные интервалы.

Результаты исследования позволят сделать выводы о текущей эффективности маркетплейса и разработать рекомендации для улучшения его работы.

### **Описание используемых данных**

#### **2.1 Источник данных**

Для анализа использованы данные из финансовой отчетности компании Ozon, доступные на официальном сайте в разделе для инвесторов[[1]](#footnote-1). Источник предоставляет “датабук” в формате Excel с ключевыми метриками, включая операционные и финансовые показатели, а также отчеты о финансовом положении и движении денежных средств.

#### **2.2 Описание данных**

Данные подразделены на следующие категории:

1. **Операционные метрики**:
   * GMV (Gross Merchandise Value) – валовый объем продаж.
   * Доля маркетплейса в GMV (%).
   * Количество заказов (млн).
   * Число активных покупателей (млн).
2. **Финансовые показатели (PnL)**:
   * Общий доход (Total revenue).
   * Прибыль/убыток за период ((Loss)/profit for the period).
   * Скорректированная EBITDA (Adjusted EBITDA).
3. **Отчет о финансовом положении**:
   * Долгосрочные активы (Total non-current assets).
   * Краткосрочные активы (Total current assets).
   * Денежные средства (Cash and cash equivalents).
   * Средства в кредитных организациях (Held by the Group’s credit institutions).
   * Общие активы (Total assets).
   * Собственный капитал (Total equity).
   * Долгосрочные обязательства (Total non-current liabilities).
   * Краткосрочные обязательства (Total current liabilities).
   * Общие обязательства (Total liabilities).
4. **Отчет о движении денежных средств**:
   * Изменения в оборотном капитале (Movements in working capital).
   * Чистые денежные потоки от операционной деятельности (Net cash (used in)/generated from operating activities).
   * Капитальные затраты (Capital expenditures).
   * Чистые денежные потоки от инвестиционной деятельности (Net cash (used in)/generated from investing activities).
   * Чистые денежные потоки от финансовой деятельности (Net cash (used in)/generated from financing activities).

Данные охватывают несколько отчетных периодов, что позволяет проводить временной анализ и выявлять динамику ключевых показателей. Перед обработкой выполнена проверка на полноту и качество данных.

### **Расчёт параметров регрессионного уравнения**

Целью данного этапа является построение регрессионной модели, которая отражает . Для анализа выбраны ключевые показатели из данных, описанных ранее.

## **3.1 Формулировка модели**

Регрессионное уравнение имеет вид:

Y=β0+β1X1+β2X2+⋯+βnXn+ε ,

где:

* Y – зависимая переменная (количество покупок за квартал);
* X1,X2,…,Xn ​ – независимые переменные (метрики, такие как GMV, доля маркетплейса, количество заказов и др.);
* β0 – свободный член (пересечение с осью Y);
* β1,β2,…,βn​ – коэффициенты регрессии, которые отражают вклад соответствующих переменных в итоговый результат;
* ε– случайная ошибка.

## **3.2 Выбор переменных**

Для построения модели в качестве зависимой переменной (Y) выбраны показатели:

* Количество заказов;

Независимые переменные (X) включают:

* GMV;
* Долю маркетплейса в GMV (%);
* Число активных покупателей.
* Долгосрочные активы
* Денежные средства
* Все активы
* Чистые денежные потоки от операционной деятельности
* Общий капитал и обязательства
* Общий капитал
* Капитальные затраты
* Чистые денежные потоки от инвестиционной деятельности
* Чистые денежные потоки от оборотного капитала

## **3.3 Процедура расчёта**

1. **Предобработка данных**:
   * Устранены пропуски в данных.
   * Приведены все показатели к сопоставимым единицам измерения.
   * Выполнена стандартизация переменных для обеспечения корректности оценки коэффициентов.
2. **Построение модели**:  
   Используется метод наименьших квадратов (МНК), реализованный с помощью библиотеки statsmodels в Python.
3. **Оценка параметров**:  
   Вычислены коэффициенты β0,β1,…,βn ​. Их значения интерпретируются с точки зрения влияния независимых переменных на результат.

Коэфициенты b при каждом члене равны соответственно:

GMV incl. services = 268.46

Number of active buyers, millions = 107.90

Total current assets = 416.99

Cash and cash equivalents = -181.02

Total assets = 10673243.42

Total equity = -42.94

Total equity and liabilities = -10673510.31

Movements in working capital1 = -50.42

Net cash (used in) / generated from operating activities1 = 72.65

Net cash (used in)/ generated from investing activities = 17.72

# **4. Анализ адекватности регрессионного уравнения**

Для оценки качества построенной модели и её применимости в прогнозировании проводится анализ адекватности регрессионного уравнения. Это включает проверку качества подгонки, проверку статистических гипотез и условий корректности оценок параметров, а также интерпретацию коэффициентов с экономической точки зрения.

## **4.1 Показатели качества подгонки**

Основные метрики, используемые для оценки качества модели:

1. **Коэффициент детерминации** :  
   Показывает долю вариации зависимой переменной, объяснённую моделью. Значение R2 близкое к 1 свидетельствует о высокой объясняющей способности модели.

R2= 0,999

1. **Скорректированный коэффициент детерминации )**:  
   Учитывает число независимых переменных, избегая переоценки качества модели при добавлении новых факторов.

R2\_adj= 0,999

## **4.2 Проверка гипотез**

1. **Значимость коэффициентов регрессии**:  
   Проверяется гипотеза H0:βi=0 против альтернативы H1:βi≠0.
   * Используется t-статистика:
   * Если p-value < 0.05, коэффициент считается значимым.

Проверка показала, что все коэффициенты значимы, что подтверждает их вклад в объяснение вариации зависимой переменной:

x1 0.00

x2 0.00

x3 0.00

x4 0.00

x5 0.04

x6 0.00

x7 0.04

x8 0.00

x9 0.00

x10 0.00

1. **Значимость уравнения в целом**:  
   Используется F-тест для проверки гипотезы H0:β1=β2=⋯=βn=0
   * Высокое значение F-статистики (p-value < 0.05) свидетельствует о значимости модели в целом.

P-value = 0.00

## **4.3 Проверка условий для получения «хороших» оценок МНК**

1. **Нормальность остатков**:  
   Проверяется с помощью теста Шапиро-Уилка.

Статистика Shapiro-Wilk: 0.9829087422813559

p-значение: 0.9428080635525465

Распределение данных похоже на нормальное

1. **Гомоскедастичность**:  
   Используется тест Спирмена. Условие равенства дисперсий выполнено, что подтверждает корректность оценок.

Корреляция Спирмена: -0.05826086956521739

p-value: 0.7868436161547347

Гетероскедастичность не обнаружена.

1. **Автокорреляция**

Статистика Дарбина-Уотсона: 2.227518413145096

Отрицательная автокорреляция. Это редко встречается в эконометрических данных, но может наблюдаться в некоторых временных рядах, где данные колеблются вокруг среднего значения.

Сгладим автокорреляцию с помощью Метода Кохрейна-Оркатта.

## **4.4 Экономический смысл коэффициентов регрессии**

Значения коэффициентов модели:

* **GMV** (β1=268.46): прирост GMV на единицу приводит к увеличению числа заказов на 268, что указывает на сильную корреляцию между валовым объемом продаж и активностью покупателей.
* **Доля маркетплейса в GMV** (β2=107.90): увеличение доли маркетплейса на 1% в составе GMV добавляет 108 заказов. Это подтверждает стратегическую важность роста доли маркетплейса для компании.
* **Число активных покупателей** (β3=416.99): прирост числа активных покупателей на 1 млн увеличивает количество заказов на 417 млн. Этот показатель подтверждает ключевую роль клиентской базы в обеспечении продаж.
* **Краткосрочные активы** (β4=−181.02): уменьшение краткосрочных активов снижает число заказов, что указывает на необходимость оптимизации оборотного капитала для поддержки операционной деятельности.
* **Капитальные затраты** (β5=72.65): рост капитальных вложений на единицу сопровождается увеличением заказов на 73. Это демонстрирует значимость инвестиций в инфраструктуру и технологии.

Общие выводы

1. **Значимость независимых переменных**: Все коэффициенты являются статистически значимыми (p-value < 0.05), что свидетельствует о их вкладе в объяснение числа заказов.
2. **Направление влияния**: Большинство коэффициентов положительны, что ожидаемо для показателей, прямо связанных с ростом операционной активности.
3. **Практическая интерпретация**: Рост GMV и числа активных покупателей оказывает наибольшее влияние на увеличение количества заказов. Эти метрики требуют приоритетного внимания в стратегическом планировании компании.

Экономический смысл коэффициентов подтверждает адекватность модели и позволяет использовать её для выработки практических рекомендаций.

# **5. Прогнозирование на основе полученной модели**

Для прогнозирования были использованы полученные регрессионные коэффициенты и независимые переменные. Прогнозируемая зависимая переменная — количество заказов.

**5.1** **Методология прогнозирования**  
Прогноз выполнен с использованием уравнения регрессии:

Для X = XСреднее \* 1.1, полученная модель выдает результат 125.6674132742919.

**5.2** **Доверительный интервал**  
Для оценки доверительного интервала использовался стандартный подход, основанный на распределении остатков. Интервалы рассчитаны с учётом стандартной ошибки прогноза и стандартного отклонения остатков.

Интервал

(120.38674611961844 – 130.94808042896537)

1. <https://ir.ozon.com/> [↑](#footnote-ref-1)