



ОТЛИЧНЫЙ ПЛАН: разработайте рекомендательную систему для создания идеальной планограммы товаров P&G

Товары не появляются на полках магазинов сами собой, как и своевременность и безошибочность в их выкладке. Человеческий фактор уходит в прошлое, настало время применять цифровые технологии и data-driven подход в прогнозировании идеальной планограммы¹. Готовы погрузиться в данные и разработать рекомендательную систему для товаров P&G? Тогда скорее читайте кейс. Успехов!

¹ Планограмма — это детальный план размещения товаров в магазине или другом торговом пространстве с целью оптимизации продаж и улучшения опыта покупателей. Подробнее в разделе «Планограмма».

ОГЛАВЛЕНИЕ



ВВЕДЕНИЕ
3 →

О КОМПАНИИ
5 →

ПЛАНОГРАММА
7 →

**ОПИСАНИЕ
ДАТАСЕТОВ**
8 →

**СБОР, ОБРАБОТКА,
АНАЛИЗ
И ХРАНЕНИЕ
ДАННЫХ**
9 →

**ТЕХНОЛОГИИ
РАСПОЗНАВАНИЯ
ФОТО**
11 →

ГЛОССАРИЙ
12 →

**КАРЬЕРНАЯ
СТРАНИЦА**
13 →

ПРИЛОЖЕНИЕ
15 →

Команда Changellenge >> подготовила этот кейс исключительно для использования в образовательных целях. Авторы не намерены иллюстрировать как эффективное, так и неэффективное решение управленческой проблемы. Кейс не содержит исчерпывающую информацию, необходимую для решения. Для построения решения вы можете использовать любые источники и свои допущения. Некоторые имена в кейсе, а также другая идентификационная информация могли быть изменены с целью соблюдения конфиденциальности.

ВВЕДЕНИЕ

Эффективная организация выкладки товаров на полках розничных магазинов играет ключевую роль в привлечении покупателей и, как следствие, увеличении продаж. Одним из основных инструментов в этом процессе является планограмма — детальная схема, определяющая расстановку товаров на полках в соответствии с их характеристиками и требованиями производителя или ретейлера.

В составлении и анализе планограммы могут применяться технологии компьютерного зрения и машинного обучения. Они используются для определения степени соответствия реальной выкладки товара на полке заданной планограмме, а также для своевременной корректировки ситуации в случае несоответствия.

Однако, сами по себе технологии не исчерпывающее решение. Прежде всего, необходимо разработать алгоритм для преобразования фотографий полок магазина с выставленными товарами в табличный вид. Входные данные состоят из фотографий, а выходные данные представляются в табличном (SQL) формате.

Но если система, собирая данные, неправильно распознает и интерпретирует товары и их расположение, это, в свою очередь, может привести к финансовым потерям, спорам с контрагентами, фактическому отсутствию товаров на полках

и снижению объемов продаж. Такие последствия подорвут доверие потребителей и контрагентов к компании, а также потенциально приведут к упущенной прибыли.

Особенно часто подобные ситуации возникают с товарами конкурентов. Важно отметить: даже если система способна безошибочно распознавать собственные товары, это не гарантирует такого же четкого и быстрого распознавания товаров конкурентов. Это связано с тем, что в режиме реального времени невозможно узнать, какие товары и в каких упаковках производят и продают конкуренты.

Также планограмма должна органично интегрироваться в существующую структуру данных. Важно тщательно продумать, как эффективно связать полученную планограмму в табличном виде с данными в уже имеющейся системе таблиц в базе данных.

Наконец, система должна не только анализировать наличие необходимых товаров и их остатки на складе магазина, последовательность выкладки, расположение позиций на первой линии, но и учитывать другие показатели. В случае выявления ошибок или потенциальных расхождений, необходимо реализовать систему оповещений для оперативного исправления ситуации.



В привычном понимании рекомендательная система — это система, которая используется для предсказания предпочтений пользователей на основе их истории взаимодействия с платформой, сервисом или товаром. Она анализирует поведение пользователей и предлагает контент на основании предположения, что может понравиться этим пользователям.

ВВЕДЕНИЕ

В кейсе на рекомендательную систему нужно посмотреть под другим углом. Как именно инструмент и его функционал будет влиять на бизнес-результаты в связке с формированием идеальной планограммы для каждого отдельно взятого магазина розничной сети? Как и какую информацию эта система будет передавать P&G и/или магазину, другим контрагентам?

Важно сфокусироваться на создании самой рекомендательной системы, а не на предоставлении рекомендации по расстановке конкретных товаров на полках. Рекомендательная система должна быть универсальна — не нужно концентрироваться на решении проблем конкретного товара или бренда из предоставленного набора данных.

ЛАЙФХАК ОТ ЭКСПЕРТА

На связи Георгий Каратеев, старший менеджер по обогащению данных. Спасибо за интерес к кейсу P&G! Хочу пожелать вам удачи, а еще поделиться 5 лайфхаками, которые помогут как решить кейс, так и сделать крутую карьеру.

Относитесь к проектам в вашей зоне ответственности, как к управлению собственным бизнесом.

Рекомендательная система сама по себе бесполезна — должно быть обоснование, как инструмент будет положительно влиять на бизнес-результаты. Нужна не просто идеальная планограмма, а система с перечнем функционала, решающего конкретные задачи, каждая из которых прямо или косвенно влияет на увеличение продаж.

Чтобы понять, с какими данными работает P&G, можно ознакомиться с набором данных, приложенным к данному кейсу.

Таким образом, для бизнеса важно иметь универсальную рекомендательную систему, которая способна генерировать идеальную планограмму с целью максимизации объема продаж.

Исходя из этой идеальной планограммы, бизнес должен получать уведомления для отдельных магазинов, указывающие на неправильное размещение товаров, а также рекомендации по внесению каких-либо изменений или принятию других решений.

ПОСТАНОВКА ЗАДАЧИ

Разработайте концепцию рекомендательной системы для товаров P&G в розничной сети магазинов с целью автоматического получения идеальной планограммы. Для этого:

1. Изучите, обработайте и проанализируйте представленные данные. Определите, какие полезные инсайты можно из них получить, чтобы повлиять на объем продаж товаров P&G.
2. Предложите эффективный подход к сбору, обработке и хранению данных планограммы в SQL-type-формате, дополните предоставленную схему данных.
3. Предложите подход к созданию рекомендательной системы: какие параметры и дополнительные данные могут быть использованы, как именно инструмент и его функционал будет влиять на увеличение продаж.

О КОМПАНИИ

Procter & Gamble — крупнейший в мире производитель товаров для дома, здоровья и личной гигиены. Продукция компании продается в 180 странах, в 70 из них есть офисы P&G. Компания Procter & Gamble была создана в 1837 году как небольшой семейный бизнес в городе Цинциннати (штат Огайо, США).

Головной офис компании в России находится в Москве, производства — в Санкт-Петербурге и Новомосковске, а сотрудников отдела продаж можно встретить во многих городах России, от Краснодара до Владивостока.

Сегодня глобальный масштаб — сильная сторона P&G, дающая каждому члену команды возможности для личностного роста, профессионального обучения и развития лидерских навыков.

Бренды P&G



О КОМПАНИИ

IT В P&G

Внутренняя инфраструктура

- техническое оснащение офиса
- внедрение систем для поддержания цепочки поставок и производства

Архитектура приложений и интеграции

- построение архитектуры облачных систем

Аналитика и работа с данными

- бизнес аналитика
- инженерия данных

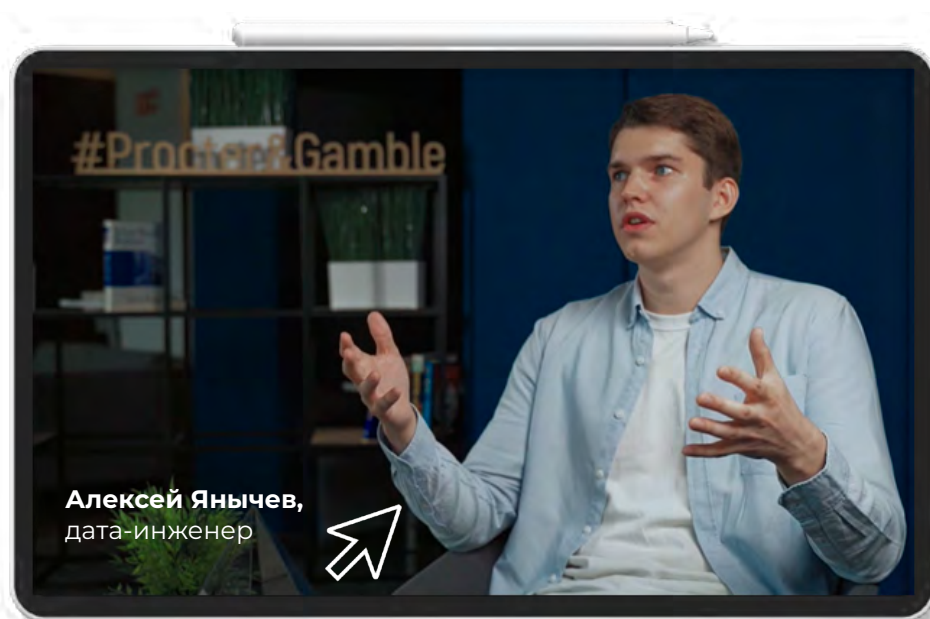
Информационная безопасность

Узнай больше про некоторые направления от наших сотрудников!



Анна Ефременко,
менеджер мастер-данных

После опыта стажировок в бизнес-акселераторе и госкорпорации я поняла, что хочу найти золотую середину: IT-направление внутри международной компании показалось мне идеальным вариантом. Так я решила пройти отбор в P&G на позицию IT Project Management Intern. Во время стажировки я тестировала поставщиков сервиса распознавания образов с помощью ИИ. Уже через 3 месяца мне предложили досрочно завершить стажировку и присоединиться к команде в качестве менеджера мастер-данных. Для перехода в новую область пришлось освоить PySpark и Power BI, погрузиться в детали процесса управления данными. На этом пути неоценимой была поддержка коллег и старших менеджеров: они всегда были готовы помочь и подсказать, в каком направлении двигаться.



Алексей Янычев,
дата-инженер

ОТЛИЧНЫЙ ПЛАН: разработайте рекомендательную систему для создания идеальной планогаммы товаров P&G



ПЛАНОГРАММА

Планограмма — это детальный план размещения товаров в магазине или другом торговом пространстве с целью оптимизации продаж и улучшения опыта покупателей. Она включает в себя информацию о том, какие товары размещаются на полках, каким должен быть их порядок, расположение и отображение, чтобы максимально эффективно использовать доступное пространство и привлечь внимание покупателей.

Эффективная планограмма может значительно увеличить продажи в магазине. Правильное размещение товаров на полках может увеличить их видимость для покупателей, стимулировать импульсивные покупки, увеличить проводимое время в магазине и повысить удовлетворенность от его посещения.

Планограммы помогают эффективно использовать пространство в магазине, учитывая такие факторы, как размеры полок, распределение света, проходимость зон, температура, влажность и другие. Планограммы часто персонализируются под конкретный магазин, разрабатываются с учетом его специфик, ассортимента, целевой аудитории и локальных особенностей.

Понятие планограммы впервые было введено в 1950-х годах, однако первые планограммы были довольно примитивными по сравнению с современными системами. С развитием технологий планограммы становятся все более автоматизированными и точными. Современные программы для создания планограмм могут использо-

вать алгоритмы машинного обучения и анализа данных для оптимизации размещения товаров.

Использование планограмм позволяет магазинам экономить время и ресурсы на ручном размещении товаров. Они облегчают процесс обновления ассортимента, сезонного изменения дизайна магазина и других операций.

ПОДХОДЫ К РАЗРАБОТКЕ ПЛАНОГРАММ

Анализ данных

Использование данных о продажах, покупательском поведении и предпочтениях потребителей для определения наиболее эффективного размещения товаров.

Моделирование и визуализация

Использование программ для моделирования и визуализации пространства магазина для создания оптимальных планограмм и оценки их эффективности.

Применение машинного обучения

Использование нейросетей и методов машинного обучения для анализа данных и прогнозирования эффективности различных вариантов планограмм, а также для распознавания объектов на фото.

Клиентские обзоры и тестирование

Получение обратной связи от клиентов и проведение тестов для оценки эффективности планограмм и внесения необходимых корректировок.



ЛАЙФХАК ОТ ЭКСПЕРТА

Фокусируйтесь на поставленных целях и задачах. Ловушка заключается в том, что ваше решение может быть креативным, но при этом не всегда решать обозначенные боли бизнеса.

ОПИСАНИЕ ДАТАСЕТОВ

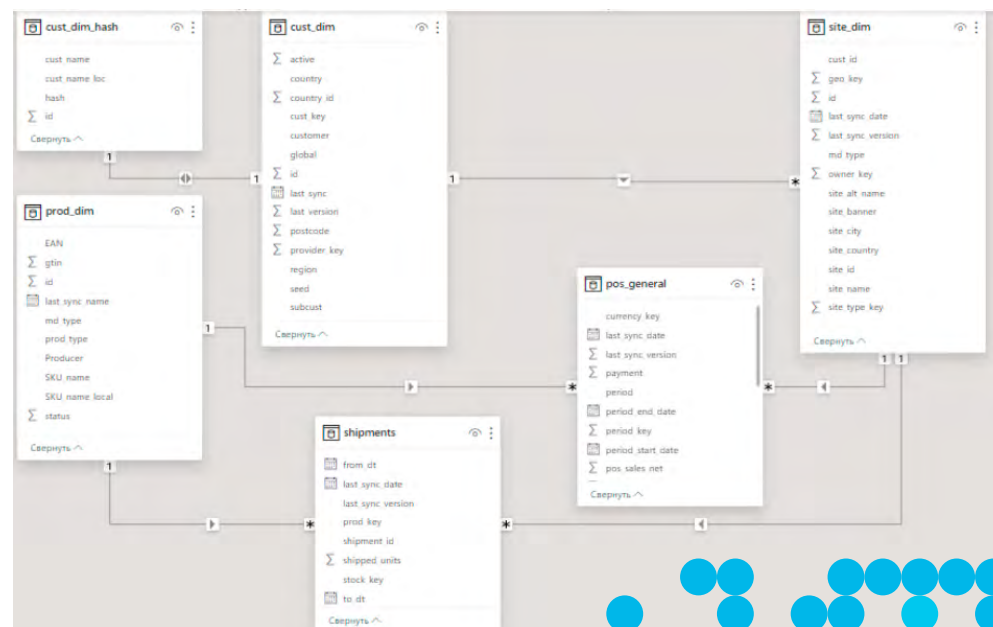
К кейсу приложены шесть датасетов в формате .csv. Ниже представлены структура данных² и общее описание датасетов, более подробное описание полей находится в приложении к кейсу.

Датасет	Тип данных	Пояснение
pos_general	FACT	Здесь хранятся фактические данные с POS ³ . Одна транзакция — одна строка (объект). Статус: в это время купили такое-то количество этого продукта по такой-то цене и такой-то остаток на полке/складе. Эти данные также дополнительно хранят количество оставшихся товаров на полке, на складе магазина. Здесь встречаются как товары P&G, так и товары конкурентов (однако по конкурентам предоставляется неполная информация)
prod_dim	DIM	Это мастер-данные по продуктам (DIM), здесь хранится перечень исключительно продуктов P&G
site_dim	DIM	Это мастер-данные по магазинам (DIM), здесь хранится перечень всех магазинов, в которые отгружаются товары P&G и каким клиентам они принадлежат
cust_dim	DIM	Это мастер-данные по клиентам (DIM), здесь хранится перечень всех клиентов P&G, однако вместо наименований клиентов хранятся зашифрованные значения в виде хешей (hash)
cust_dim_hash	DIM	Вспомогательная расшифровка хешей для клиентов. В данном случае тут деанонимизирован один условный клиент
shipments	FACT	Здесь хранятся данные по отгрузкам. В какое время какое количество продуктов отгрузили в конкретный магазин

ЛАЙФХАК ОТ ЭКСПЕРТА

«Less Efforts — More Benefits»: если вы понимаете, что для решения проблемы бизнеса необходим сложный, дорогой и комплексный проект, начните с шагов, которые потребуют минимальных усилий и принесут больше выгоды бизнесу.

Структура базы данных



² Связь «1-*» — one2many, Связь «1-1» — one2one

³ Points of sales — точка продаж, кассовые аппараты.

СБОР, ОБРАБОТКА, АНАЛИЗ И ХРАНЕНИЕ ДАННЫХ

Для сбора, обработки и хранения данных текущей планограммы магазина можно использовать ряд современных технологий и инструментов.

Интеграция с POS-системами для сбора данных

Для получения актуальных данных о продажах и наличии товаров в магазине можно использовать данные с POS-систем магазинов. Это может позволить автоматически получать данные о продажах и остатках товаров на складе, что может помочь оптимизировать планограмму и улучшить управление запасами.

Также для сбора данных могут быть использованы технологии распознавания фото и компьютерного зрения. Подробнее о них — в разделе «Технологии распознавания фото».

Python для обработки и анализа данных

Универсальный, высокоуровневый язык программирования, который отличается чистым синтаксисом и простотой в изучении. Его богатая стандартная библиотека и активное сообщество разработчиков делают Python идеальным выбором для широкого спектра задач — от анализа данных до веб-разработки.

Описание библиотек

Pandas	Библиотека для работы с данными в форме таблиц (или DataFrame). Она предоставляет инструменты для чтения, записи, фильтрации, преобразования и агрегации данных, делая процесс анализа данных быстрым и эффективным
NumPy	Библиотека для работы с многомерными массивами и выполнения математических операций над ними. С его помощью можно выполнять широкий спектр операций — от простых арифметических действий до сложных операций линейной алгебры
Matplotlib	Библиотека для создания высококачественных графиков и визуализации данных. С ее помощью можно создавать графики, диаграммы, гистограммы и многое другое, что делает визуализацию данных простой и эффективной
Seaborn	Библиотека для визуализации данных, построенная поверх Matplotlib, которая облегчает создание красивых и информативных графиков. Она предоставляет высокоуровневые интерфейсы для создания различных типов графиков и улучшает их внешний вид по умолчанию
Scikit-learn	Библиотека машинного обучения с открытым исходным кодом, предоставляющая простые и эффективные инструменты для анализа данных и построения моделей машинного обучения. С ее помощью можно создавать и обучать модели для решения различных задач, таких как классификация, регрессия, кластеризация и другие
TensorFlow и PyTorch	Фреймворки глубокого обучения, разработанные для создания и обучения нейронных сетей на больших объемах данных. Оба фреймворка активно используются для решения различных задач машинного обучения и искусственного интеллекта
OpenCV	Предоставляет широкий набор инструментов для обработки изображений, включая функции для детектирования объектов, сегментации изображений, распознавания образов и многое другое

СБОР, ОБРАБОТКА, АНАЛИЗ И ХРАНЕНИЕ ДАННЫХ

BI-инструменты для анализа данных

Power BI от Microsoft — мощный инструмент для визуализации и анализа данных, который позволяет объединять данные из различных источников, создавать интерактивные дашборды и отчеты, а также интегрироваться с другими продуктами Microsoft.

Tableau Public от Tableau Software предоставляет возможность создавать и публиковать интерактивные визуализации данных онлайн. Пользователи могут подключаться к различным источникам данных, создавать разнообразные визуализации и обмениваться результатами работы с другими пользователями.

ЛАЙФХАК ОТ ЭКСПЕРТА

Помните про work-life balance: планируйте время не только под работу, но и под отдых. Выгорание негативно скажется на Вашей эффективности.

SQL или NoSQL для хранения данных

Для хранения данных планogramмы магазина можно использовать базу данных (БД) SQL, такую как PostgreSQL или MySQL, или NoSQL, такую как MongoDB или Cassandra. БД SQL обеспечивают структурированное хранение данных с возможностью выполнения сложных запросов, тогда как БД NoSQL предоставляют гибкую схему данных для хранения неструктурированных данных. MySQL и PostgreSQL относятся к реляционным БД. Они организуют данные в виде таблиц, состоящих из рядов и столбцов, где каждый ряд представляет собой запись, а каждый столбец — атрибут. Они используют структурированный язык запросов (SQL) для выполнения операций.

MongoDB, Cassandra относятся к нереляционным БД. Они отличаются тем, что не требуют фиксированной схемы и могут хранить данные в различных форматах, таких как документы, графы, столбцы или ключ-значение. Они часто используются для работы с большими объемами неструктурированных данных и могут обеспечить более гибкое масштабирование.



ТЕХНОЛОГИИ РАСПОЗНАВАНИЯ ФОТО

Распознавание объектов на фотографиях и их преобразование в табулярные данные — важное направление в области компьютерного зрения и анализа изображений.

Процесс можно разделить на несколько этапов:

1. Обнаружение объектов на фотографии

С помощью различных алгоритмов и нейросетевых моделей происходит выделение и обнаружение объектов на изображении. Например, объекты могут быть выделены с помощью методов обнаружения объектов, таких как метод каскадных классификаторов Хаара, методы на основе признаков Хогга или современные нейронные сети, такие как Faster R-CNN, YOLO или SSD.

2. Классификация объектов

После обнаружения объектов происходит их классификация, то есть определение, к какому классу или категории каждый объект принадлежит. Для этого также используются различные алгоритмы машинного обучения и нейросетевые модели, обученные на больших наборах размеченных изображений.

3. Преобразование в табличный формат

Полученные результаты обнаружения и классификации объектов могут быть преобразованы в табличный формат путем записи различных характеристик объектов в виде строк и столбцов таблицы. Например, для каждого обнаруженного объекта можно записать его координаты на изображении, класс или категорию, уверенность модели в правильности классификации и другие характеристики.

4. Анализ и использование табличных данных

Полученные табличные данные могут быть использованы для различных целей анализа, в том числе для статистического анализа объектов на изображении, их распределения по категориям, а также для дальнейшего принятия решений, например, в области мониторинга трафика на дорогах, анализа медицинских изображений или управления инфраструктурой городов.



ЛАЙФХАК ОТ ЭКСПЕРТА

Экспериментируйте, мыслите нестандартно, но не забывайте про изучение опыта ваших коллег и других компаний на рынке.

ГЛОССАРИЙ

SQL-type

Формат данных, который может быть сохранен или извлечен из реляционной базы данных с использованием языка структурированных запросов (SQL). Такие типы данных включают в себя целочисленные значения, десятичные числа, строки, даты и времена, булевы значения и другие. Они определяют формат, в котором данные будут храниться в таблицах базы данных и как они будут обрабатываться при выполнении запросов SQL.

Табулярные данные

В контексте обработанных фотографий представляют собой структурированный формат данных, который описывает информацию о наличии товаров на полках магазина на основе обработанных фотографий. Эти данные могут включать в себя идентификаторы товаров, количество товаров на полке, расположение товаров и другие характеристики. Табулярный формат удобен для хранения и анализа данных, а также для интеграции с другими системами и инструментами аналитики.

Метрики On-Shelf Availability и Out-of-Stock

Метрика OSA (On-Shelf Availability) отражает процентное соотношение товаров, которые доступны для покупки на полках магазина в определенный момент времени.

Стремление к высокому уровню OSA помогает предотвращать потерю продаж из-за недостатка товаров на полках.

Метрика OOS (Out-of-Stock) указывает на процентное соотношение времени, в течение которого товар отсутствует на полках магазина или недоступен для покупки. Высокий уровень OOS может привести к потере продаж и недовольству покупателей и потере доверия/лояльности к бренду, производителю.

Метрики OSA и OOS можно рассчитывать как для всех товаров, так и для товаров отдельного бренда или даже отдельного SKU. Некоторые компании придумывают свои производные метрики от этих и других данных, чтоб получить свои более полезные и информативные метрики.

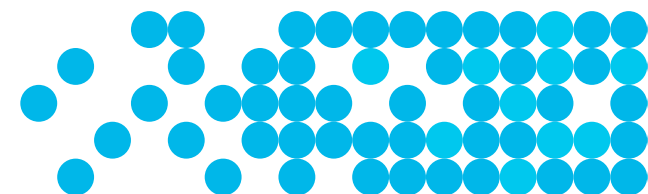
Другие связанные метрики могут включать в себя:

- Sell-Through Rate. Измеряет процент товаров, которые были проданы в отношении общего количества товаров, поставленных в магазин за определенный период.
- Rate of Sale. Отражает скорость, с которой товары продавались за определенный период, и может помочь определить их популярность и востребованность.
- Inventory Turnover. Метрика показывает, как часто запасы товаров обновляются и перепродаются за определенный период, что помогает оптимизировать управление запасами.

Virtual Stock

В контексте кейса информация в системе о наличии товаров на складах и полках магазинов, но его фактическое отсутствие на полках, следствием чего может являться отсутствие продаж. Причины могут быть разными (потеря/порча товара, кража), но эффект один — снижение объемов продаж. Новые товары не доставляются на склад и не выставляются на полки, так как, согласно учетным системам, товар имеется на полке и на складах. В таком случае важно своевременно в кратчайшие сроки обнаружить «виртуальные стоки» и затем их списать, чтобы возобновились поставки товара в магазин и он выставился на полку.

Необходимо учитывать, что время обнаружения «виртуального стока» — это время, в течение которого компания теряет потребителей и их лояльность. В краткосрочной перспективе может выражаться в упущенной прибыли, а в долгосрочной — в снижении объема продаж.



СТАЖИРОВКИ В P&G

Стажировка в P&G – это работа над полноценным проектом, который ты будешь лидировать с первого дня в компании. Мы ищем талантливых и амбициозных студентов и выпускников, которые хотят воплощать в жизнь смелые идеи, искать креативные пути решения бизнес-задач и менять жизни миллионов людей к лучшему!

ТЕБЯ ЖДУТ:

- Масштабные проекты и реальные бизнес-задачи
- Возможность поработать со всемирно известными брендами
- Международный опыт в одной из крупнейших FMCG-компаний мира
- Возможность получить оффер на постоянную позицию в случае успешного результата стажировки

А ТАКЖЕ:

- Зарботная плата 90 000 рублей в месяц до вычета НДФЛ на протяжении всего периода стажировки
- Ежемесячное субсидирование расходов на питание
- Гибкий режим рабочего времени и гибридный формат работы
- Доступ к глобальным обучающим ресурсам P&G
- Доступ к корпоративному провайдеру скидок и привилегий PrimeZone

ТРЕБОВАНИЯ:



Студенты 3+ курса бакалавриата или недавние выпускники



Уровень английского Upper-Intermediate и выше



Возможность работать 40 часов в неделю

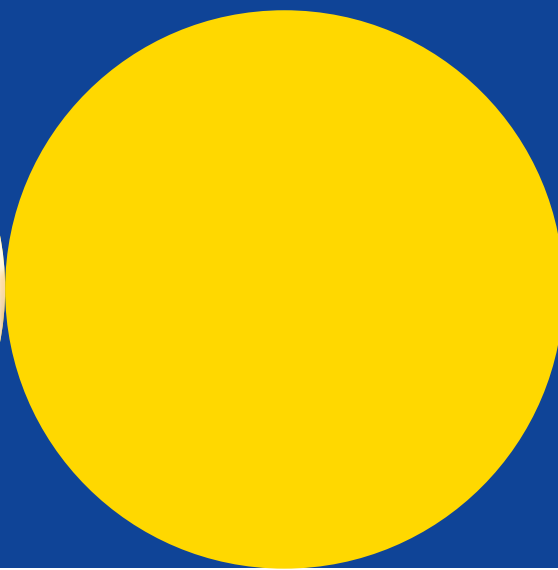
Набор на стажировки ведется круглый год. Актуальные вакансии можно найти и откликнуться на них по [ссылке](#).

В НАСТОЯЩЕЕ ВРЕМЯ В IT-КОМАНДЕ P&G ОТКРЫТЫ СЛЕДУЮЩИЕ ВАКАНСИИ:

[Стажер дата-инженер](#)

[Стажер отдела информационной безопасности](#)

[Стажер-менеджер проектов](#)



ЭТАПЫ ОТБОРА В P&G

1. Выбор вакансии

Прежде чем откликнуться на вакансию, внимательно прочтите ее описание. Если описание не дает достаточного представления о том, чем предстоит заниматься на практике, ознакомьтесь с карьерными историями успешных стажеров в сообществе P&G в VK: <https://vk.com/pgrossia>. Актуальные вакансии можно найти и откликнуться на них по [ссылке](#).

2. Подача заявки

После заполнения небольшой анкеты загрузите короткое (лучше одностраничное) резюме на английском языке. Сопроводительное письмо (cover letter) прикладывать необязательно.

3. Онлайн-тестирование

Если ваши ответы на вопросы анкеты подходят P&G, после ее заполнения вам будет предложено пройти онлайн-тестирование. Мы рекомендуем проходить его в спокойной обстановке с персонального компьютера.

Тестирование состоит из двух частей:

- Первая часть, Peak Performance Assessment, не ограничена по времени и направлена на оценку вашей культурной совместимости с компанией. Узнать подробнее об особенностях тестирования вы можете на сайте <https://www.pgcareers.com/global/en/peak-performance-assessment>.
- Вторая часть, Interactive Assessment, ограничена по времени и направлена на оценку когнитивных способностей. Она требует большего фокуса, поэтому, прежде чем начать ее выполнение, убедитесь, что вас ничто не отвлекает. Узнать больше об особенностях тестирования вы можете на сайте <https://www.pgcareers.com/global/en/interactive-assessments>.

4. Серия онлайн-интервью

В случае успешного прохождения онлайн-тестирования вас пригласят на серию из трех интервью с сотрудниками компании. Интервью проводят действующие менеджеры P&G, поэтому воспользуйтесь возможностью спросить обо всех деталях, которые вас интересуют.

Лучший способ подготовиться — вспомнить и подумать о личных, академических и/или профессиональных достижениях, которыми вы гордитесь. Подробнее об интервью в P&G можно узнать на сайте <https://www.pgcareers.com/hiring-process#interviews-section>.

5. Предложение о работе

В течение недели с момента получения письма об успешном прохождении финального интервью с вами свяжется ваш потенциальный руководитель или HR-партнер для обсуждения деталей оффера. Если с вами не связались в течение недели, напишите об этом на почту pgeuropcareers.im@pg.com.



ПРИЛОЖЕНИЕ 1.

ОПИСАНИЕ ПОЛЕЙ ДАТАСЕТОВ

pos-general

Поле	Первичный ключ	Вторичный ключ	Пояснение
site_key	No	Yes: site_dim.site_id	ID магазина, из которого пришла информация
last_sync_date	No	No	Время последней синхронизации данных с источником
prod_key	No	Yes: prod_dim.EAN	ID продукта, по которому сделана запись
currency_key	No	No	Валюта, в которой была совершена транзакция
period_key	No	No	Ключ периода, в нашем случае период. Могут быть разные периоды: 3 минуты, 5 минут, час, день, неделя, месяц. У каждого такого периода свой ключ
period	No	No	Период, за который указана транзакция
source	No	No	Источник данных
period_start_date	No	No	Старт периода
period_end_date	No	No	Конец периода
promo	No	No	Есть ли текущая промоактивность. True — да, False — нет
payment	No	No	Оплата картой или наличными: 0 — наличные, 1 — безналичная оплата
pos_sales_qty	No	No	Количество проданного товара
pos_sales_with_tax	No	No	Сумма продажи с налогом
pos_sales_net	No	No	Сумма продажи без налога
pos_store_qty	No	No	Количество магазинов в кластере, к которому прикреплен данный site_key
units_on_hand	No	No	Количество товара на полке в данный момент
units_on_stock	No	No	Количество товара на складе магазина в данный момент
tax	No	No	Налог

ПРИЛОЖЕНИЕ 1.

ОПИСАНИЕ ПОЛЕЙ ДАТАСЕТОВ

prod_dim

Поле	Первичный ключ	Вторичный ключ	Пояснение
id	Yes	No	ID продукта
SKU_name	No	No	Наименование продукта
SKU_name_local	No	No	Локальное наименование продукта
last_sync_name	No	No	Время последней синхронизации данных с источником
EAN	Yes	No	EAN-код продукта
Producer	No	No	Производитель продукта
md_type	No	No	Тип мастер данных
gtin	No	No	GTIN-код продукта
prod_type	No	No	Категория продукта
status	No	No	Статус продукта: 0 — продукт еще не вышел на рынок, 1 — продукт на рынке, 2 — продукт снят с рынка, 3 — иное

ПРИЛОЖЕНИЕ 1.

ОПИСАНИЕ ПОЛЕЙ ДАТАСЕТОВ

site_dim

Поле	Первичный ключ	Вторичный ключ	Пояснение
id	Yes	No	ID магазина (канала) — скорее index
site_id	Yes	No	ID канала
cust_id	No	Yes: cust_dim.cust_key	ID клиента
last_sync_version	No	No	Версия объекта — информации по каналу
last_sync_date	No	No	Дата последней синхронизации данных
md_type	No	No	Тип мастер-данных
site_type_key	No	No	Тип магазина: 1 — крупный тип (супермаркеты), 2 — магазины у дома
owner_key	No	Yes: cust_dim.cust_key	ID владельца канала
geo_key	No	No	Ключ региона
site_name	No	No	Hash наименования канала
site_alt_name	No	No	Hash альтернативного наименования канала
site_banner	No	No	Обозначение канала
site_city	No	No	Локация канала
site_country	No	No	Страна канала

ПРИЛОЖЕНИЕ 1.

ОПИСАНИЕ ПОЛЕЙ ДАТАСЕТОВ

cust_dim

Поле	Первичный ключ	Вторичный ключ	Пояснение
id	Yes	No	ID клиента — скорее index
provider_key	No	No	Ключ источника данных. Это условно может быть OLTP-система
cust_key	Yes	No	Ключ (ID) клиента
last_version	No	No	Версия объекта — информации по клиенту
last_sync	No	No	Дата последней синхронизации данных
region	No	No	Регион клиента
global	No	No	Клиент указан в глобальной или только локальных системах
seed	No	No	Обозначение региона
active	No	No	Клиент активный: 1 — да, 2 — нет
customer	Yes	Yes: cust_dim_hash.hash	Hash имени клиента
subcust	No	No	Вторичный клиент
postcode	No	No	Индекс клиента
country	No	No	Страна
country_id	No	No	ID страны

ПРИЛОЖЕНИЕ 1.

ОПИСАНИЕ ПОЛЕЙ ДАТАСЕТОВ

cust_dim_hash

Поле	Первичный ключ	Вторичный ключ	Пояснение
id	Yes	No	ID расшифровки — скорее index
hash	Yes	Yes: cust_dim.customer	Hash зашифрованного клиента
cust_name_loc	No	No	Локальное наименование клиента
cust_name	No	No	Глобальное наименование клиента

shipments

Поле	Первичный ключ	Вторичный ключ	Пояснение
last_sync_date	No	No	Дата последней синхронизации данных
last_sync_version	No	No	Версия объекта — информации по отгрузке
prod_key	No	Yes: prod_dim.EAN	Ключ отгруженного продукта
shipment_id	Yes	No	Уникальный ID отгрузки в виде хеша
stock_key	No	Yes: site_dim.site_id	Ключ места отгрузки (канала)
shipped_units	No	No	Количество отгруженных товаров
from_dt	No	No	Timestamp начала периода
to_dt	No	No	Timestamp конца периода

CHANGELLENGE >>

Кейс написан и опубликован
Changellenge >> —
ведущей организацией
по кейсам в России.

www.changellenge.com
info@changellenge.com
vk.com/changellengeglobal



Кейс создан по заказу
Procter & Gamble