

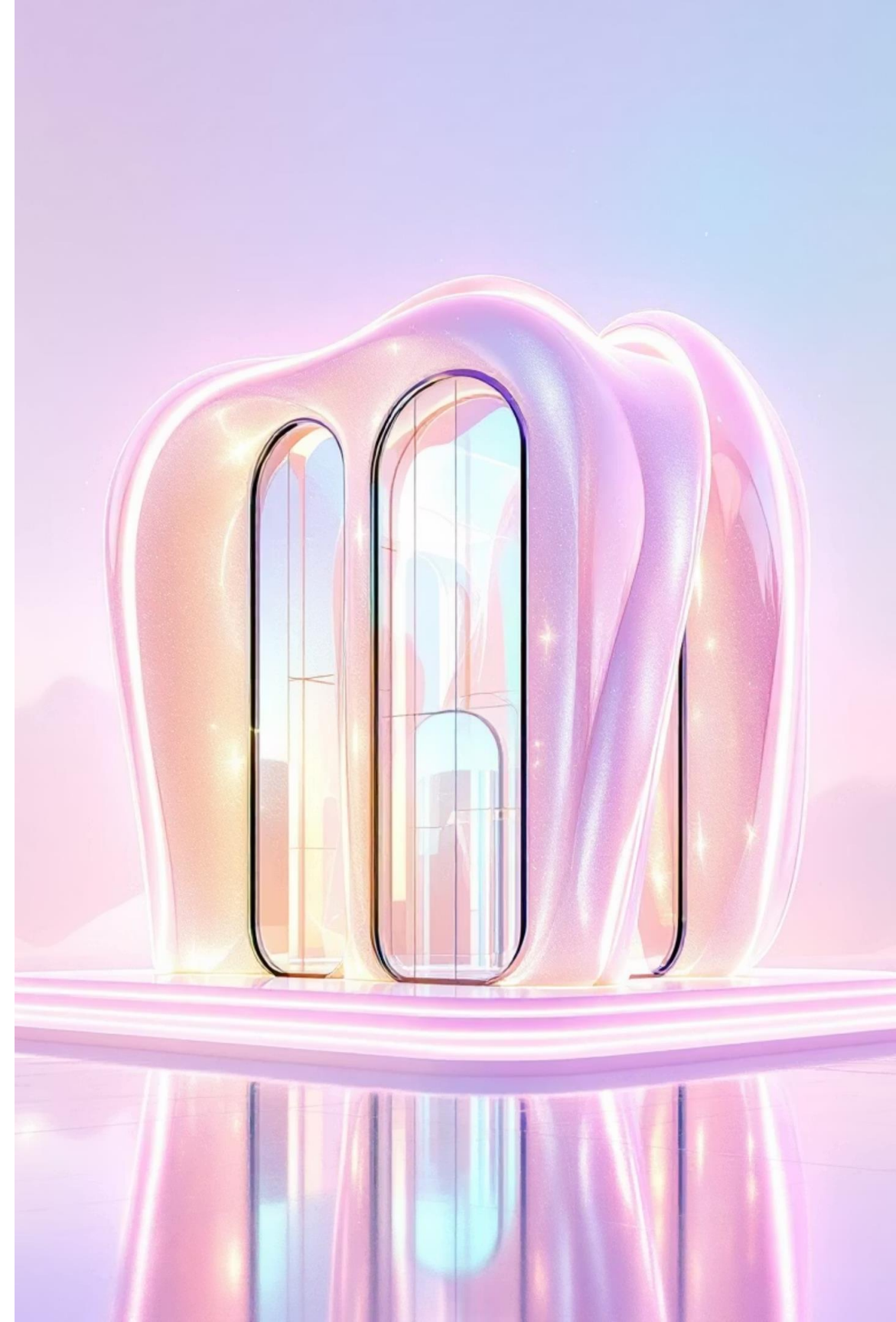
Анализ клиентов банка

Исследование поведения и факторов принятия решения о вкладе

Выполнил: Сингаевский Андрей Николаевич

Дата: Сентябрь 2025

Источник данных: Kaggle — Bank Marketing Dataset



Цель исследования и описание данных

Цель исследования

Выявить ключевые характеристики клиентов, влияющие на решение оформить срочный вклад (term deposit). Данное исследование направлено на глубокое понимание поведенческих паттернов клиентов банка и выявление факторов, которые наиболее эффективно способствуют конверсии в рамках маркетинговых кампаний.

Анализ позволит оптимизировать стратегию привлечения клиентов к депозитным продуктам, сократить затраты на неэффективные коммуникации и повысить общую результативность банковских маркетинговых активностей.


Датасет содержит информацию о клиентах португальского банка и результатах телефонных маркетинговых кампаний, проводившихся с целью привлечения клиентов к открытию срочных депозитов. Каждая запись содержит демографические данные клиента, историю взаимодействия с банком и итоговый результат кампании.

О датасете

- **11 162 записи** — репрезентативная выборка клиентов
- **17 признаков:** возраст, профессия, образование, баланс, тип контакта, длительность звонка и другие
- **Целевая переменная:** deposit (yes/no)
- Данные собраны в ходе прямой маркетинговой кампании по телефону

Этапы предобработки данных

01	02	03
<div>Приведение названий столбцов к стилю camelCase</div> <div>Обеспечение единообразия и соответствия требованиям технического задания. Все названия столбцов преобразованы в удобный для работы формат, что упрощает дальнейший анализ и минимизирует риск ошибок при обращении к данным.</div>	<div>Преобразование категориальных признаков</div> <div>Категориальные переменные, такие как профессия (job), образование (education), семейное положение (marital), были проанализированы и подготовлены для дальнейшего статистического анализа. Проведена кодировка и группировка редких категорий.</div>	<div>Проверка на дубликаты</div> <div>Выполнена тщательная проверка датасета на наличие дублирующихся записей. Результат: дублей не обнаружено, что подтверждает высокое качество исходных данных и их готовность к анализу.</div>
04	05	
<div>Анализ пропусков</div> <div>Проведена детальная проверка всех столбцов на наличие пропущенных значений. Результат: пропущенных значений не обнаружено, датасет является полным и не требует импутации.</div>	<div>Преобразование временных признаков</div> <div>Признак длительности звонка (duration) и дата последнего контакта преобразованы в удобный для анализа формат. Созданы дополнительные производные признаки для более глубокого исследования временных паттернов.</div>	

 **Комментарий из кода:** «Приводим к camelCase для единообразия и соответствия требованиям ТЗ»

Качественная предобработка данных является критически важным этапом любого аналитического исследования. В данном случае высокое качество исходного датасета позволило сосредоточиться на содержательном анализе, минимизировав время на очистку данных.

Распределение клиентов по профессиям

Лидирующие категории

- **Административный персонал** — наибольшая доля в выборке, представляют офисных работников среднего звена
- **Технические специалисты** — инженеры, IT-специалисты, техники различных профилей
- **Управленцы** — менеджеры среднего и высшего звена, руководители подразделений

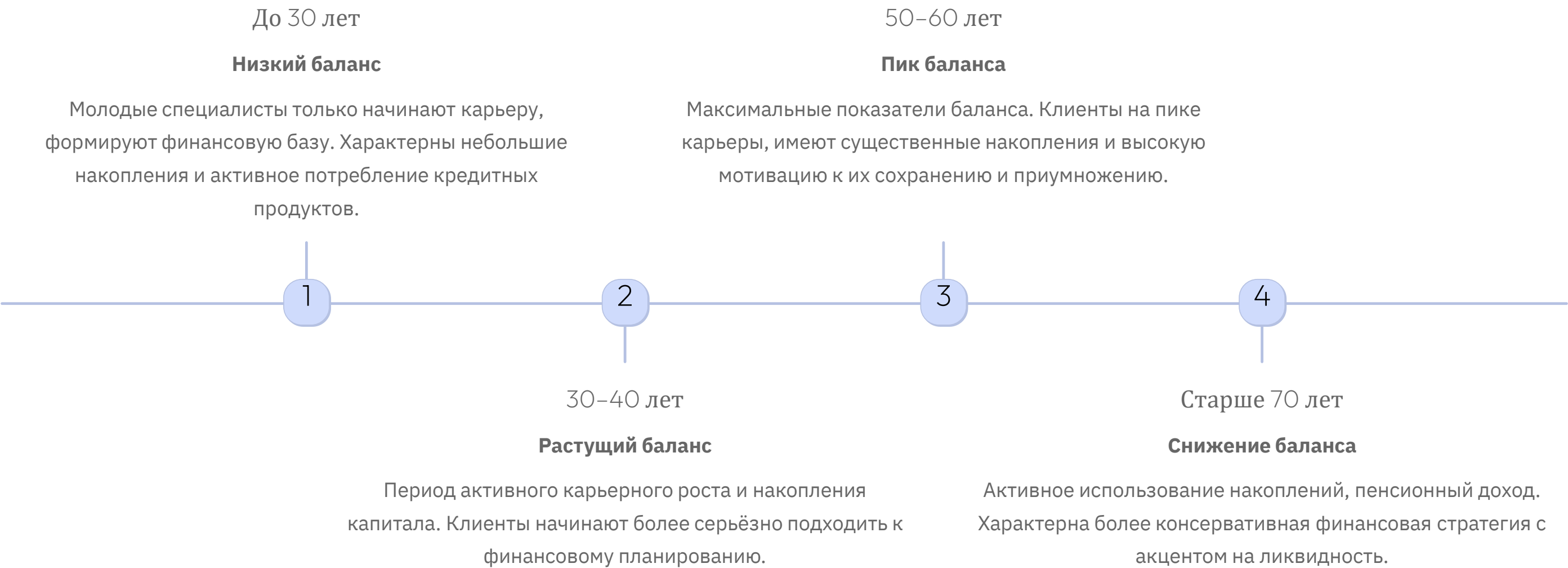
Минимальное представительство

- **Студенты** — ограниченная финансовая база, нестабильный доход
- **Пенсионеры** — консервативный подход к финансам, низкая активность
- **Предприниматели** — небольшая доля в общей базе клиентов

Профессиональная структура клиентской базы демонстрирует типичное распределение для европейского банка: преобладают представители офисных профессий и специалисты с регулярным доходом. Эти категории клиентов обладают более предсказуемыми финансовыми потоками и, как следствие, представляют особый интерес для депозитных продуктов.

Интересно отметить, что технические специалисты и управленцы, несмотря на более высокий доход, не всегда демонстрируют пропорционально высокую активность в открытии депозитов. Это может быть связано с наличием альтернативных инвестиционных стратегий или более сложных финансовых портфелей.

Средний баланс по возрастным группам



Возрастная динамика балансов демонстрирует классическую кривую жизненного цикла накоплений. Клиенты возрастной группы 50–60 лет представляют особый интерес для депозитных продуктов: они обладают максимальным капиталом и заинтересованы в его надёжном размещении с минимальными рисками.

Молодая аудитория (до 30 лет), несмотря на низкие текущие балансы, представляет стратегический интерес для банка как потенциально лояльная клиентская база на долгосрочную перспективу. Инвестиции в привлечение и удержание этой категории могут окупиться в течение следующих 10–20 лет.

Уровень подписки на депозит и длительность звонка



Ключевая корреляция

Чем дольше разговор — тем выше вероятность согласия

Длительность телефонного разговора оказалась одним из наиболее сильных предикторов успешного завершения сделки. Это не означает причинно-следственной связи, но указывает на качество диалога.

400

Медианная длительность

Средняя длительность успешного звонка составляет около 400 секунд (6–7 минут), что указывает

85%

Конверсия при долгом звонке

Звонки длительностью более 10 минут демонстрируют конверсию до 85%, что существенно выше

3X

Разница в эффективности

Разговоры длительностью более 5 минут в 3 раза более эффективны по сравнению с короткими

Главные инсайты исследования



Оптимальная аудитория: 30–50 лет

Клиенты возрастной группы 30–50 лет демонстрируют наибольшую восприимчивость к предложениям о депозитах. Это «золотая середина» между финансовой стабильностью и готовностью к диалогу с банком.



Образование имеет значение

Клиенты с высшим образованием демонстрируют статистически значимо более высокую склонность к оформлению депозитов. Это связано как с уровнем дохода, так и с финансовой грамотностью.



Качество диалога критично

Длительность разговора является сильным предиктором успеха, но важно понимать: это следствие качественного диалога, а не его причина. Содержательная беседа естественным образом требует времени.

Сезонный фактор

Клиенты, с которыми устанавливался контакт в мае–июне, демонстрировали значительно более низкую конверсию по сравнению с другими месяцами. Возможные причины: отпускной период, налоговая нагрузка в начале лета, сезонные расходы.

Баланс — не главный фактор

Вопреки ожиданиям, размер текущего баланса сам по себе слабо коррелирует с решением об открытии депозита. Гораздо важнее контекст коммуникации, доверие к банку и понимание преимуществ продукта.

Эти инсайты формируют основу для пересмотра маркетинговой стратегии банка и оптимизации работы с клиентами. Понимание ключевых факторов успеха позволяет более эффективно распределять ресурсы и фокусироваться на наиболее перспективных сегментах аудитории.

Рекомендации для маркетинга



Стратегические приоритеты

На основе проведённого анализа сформулированы конкретные рекомендации для оптимизации маркетинговых кампаний и повышения их эффективности.



Фокус на целевой аудитории

Сконцентрировать усилия на клиентах возрастной группы 30–50 лет с высшим образованием

Оптимизация скриптов

Разработать и внедрить улучшенные скрипты разговоров, направленные на увеличение качества

Сезонное планирование

Пересмотреть график кампаний: избегать массовых обзвонov в мае–июне, когда конверсия

Следующие шаги аналитики



Построение предиктивной модели

Разработать модель машинного обучения для предсказания вероятности согласия клиента на открытие депозита. Рассмотреть логистическую регрессию, деревья решений и ансамблевые методы (Random Forest, XGBoost) для достижения максимальной точности прогноза.



Feature Engineering

Создать производные признаки, которые могут улучшить качество модели: соотношение баланса к возрасту, индекс «теплоты» клиента на основе истории взаимодействий, категоризация клиентов по группам риска и потенциала.

Краткосрочная перспектива (1–3 месяца)

- Запуск A/B тестов на основе полученных инсайтов
- Обучение колл-центра новым скриптам
- Пилотный запуск предиктивной модели



Исследование взаимодействия признаков

Провести углублённый анализ взаимодействия между признаками: «профессия × баланс», «возраст × образование», «месяц контакта × длительность звонка». Выявить синергетические эффекты, которые могут не проявляться при изолированном рассмотрении факторов.



Внедрение системы мониторинга

Разработать дашборд для отслеживания ключевых метрик в реальном времени: текущая конверсия по сегментам, эффективность операторов, динамика по каналам коммуникации. Автоматизировать генерацию отчётов для менеджмента.

Долгосрочная перспектива (6–12 месяцев)

- Полномасштабное внедрение ML-модели
- Интеграция с CRM-системой банка
- Расширение анализа на другие продукты

Ссылки и контакты



Материалы исследования

Полный код анализа, визуализации и подробные комментарии доступны в публичном репозитории. Исследование выполнено с использованием Python, pandas, matplotlib, seaborn и других инструментов современной аналитики данных.



Исходный код

GitHub репозиторий:

portfolio_Python

Файл: notebooks/bank_clients_research.ipynb

https://github.com/singaevsky/portfolio_Python.git



Датасет

Источник данных:

Kaggle — Bank Marketing Dataset

Содержит 11 162 записи о клиентах португальского банка и результатах маркетинговых кампаний



Контакты

Автор исследования:

Сингаевский Андрей Николаевич

singaevsky@yandex.ru

Открыт для вопросов, обсуждения результатов и сотрудничества