

# ПРИМЕНЕНИЕ ЦИФРОВОГО ПРОГРАММИРОВАНИЯ В АУДИТЕ

## APPLICATION OF DIGITAL PROGRAMMING IN AUDIT

**Емельченко Елена Сергеевна**, студентка 4 курса бакалавриата, ВШГА МГУ  
им. М.В. Ломоносова, г. Москва

**Emelchenko E.S.**, mayer.alena2016@yandex.ru

### **Аннотация**

Целью исследования стало изучение возможностей применения цифрового программирования в сфере аудиторской деятельности и выявление современных методов оптимизации данного процесса. В этой связи было определено актуальность данной сферы, преимущества и возможности на примере компании Enron.

### **Annotation**

The purpose of the study was to study programming in the field of auditing and to identify modern methods of optimizing this process. In this regard, the relevance of this area, advantages and opportunities were determined by the example of Enron.

**Ключевые слова:** программирование, Python, data science, аудит, мошенничество, проверка, выборка, Enron.

**Keywords:** programming, Python, data science, audit, fraud, verification, sampling, Enron.

### **1. Актуальность информатизации процесса аудиторской проверки.**

Язык Python прост для изучения и использования. Он универсальный и используется в разных отраслях деятельности такие как Data Science, финансы, аудит соответствия и эффективности, разработка программного обеспечения

и другое. Любые данные, большие они или маленькие, несут огромную информацию во всех сферах использования.

Аудит - независимая проверка бухгалтерской (финансовой) отчетности аудируемого лица в целях выражения мнения о достоверности такой отчетности.<sup>1</sup> Когда речь заходит о продвижении компании по лестнице успеха, то аудиторы, финансисты и бухгалтеры не могут позволить себе допускать слишком много ошибок, которые нужно исправлять. Существует множество методов проведения аудита, и лишь в нескольких случаях за последние годы мы замечали то, как используемые технологии внедрялись вместе с процессами. Сегодня мы рассмотрим Python и то, какую роль он играет в области аудита.

Когда мы говорим о финансовой или бухгалтерской отчетности, то невольно вспоминаем огромные массивы данных. Программирование упрощает момент изъятия, формирования и выгрузки данных, а также упрощает процесс визуализации данных и позволяет проводить сложный статистический анализ. Решения на базе Python оснащены мощными алгоритмами машинного обучения, которые обеспечивают прогностическую аналитику, что чрезвычайно важно для всех пользователей финансовых услуг.

Язык Python один из самых востребованных и мобильных языков программирования, он выделяется своей скоростью разработки, простым и понятным синтаксисом, легкостью в изучении, множеством фреймворков и библиотек.

Конечно, переход с традиционных способов аудиторской проверки на цифровой представляется чрезмерно запугивающим и малоизученным, но с

---

<sup>1</sup> Федеральный закон от 01.07.2010 № 136-ФЗ, от 01.12.2014 № 403-ФЗ, от 02.07.2021 № 359-ФЗ "Об аудиторской деятельности"

иной стороны перед нами обнаруживается большой потенциал информационных технологий.

Во время пандемии Ковид-19 мы все столкнулись с проблемой логистики, когда были ограничены какие-либо встречи, выездные проверки, доступы к сайтам компаний и другие проблемы. Аудиторы были в поисках других методов сборов и проверки отчетности для достаточного количества информации, чтобы в дальнейшем осуществить аудиторское заключение.

IAASB определяет аналитику данных для аудита как науку и искусство обнаружения и анализа закономерностей, отклонений и несоответствий, а также извлечения другой полезной информации из данных, лежащих в основе аудита или относящихся к предмету аудита, посредством анализа, моделирования и визуализации с целью планирования и проведения аудита.<sup>2</sup>

Крупные зарубежные аудиторские фирмы уже используют аналитику данных как часть своих аудиторских предложений и решений, чтобы снизить риск и повысить ценность своих услуг для клиента. Не существует еще единого универсального инструмента анализа аудиторских данных, но существует множество форм и проектов, разработанных программистами-аудиторами самостоятельно.

Они используют визуальные методы, такие как графики, для представления данных, помогающих выявить тенденции и корреляции, и математические расчеты, которые помогают увидеть взаимосвязь между различными ситуациями.

Для аудиторов основной движущей силой применения анализа данных является увеличение качества аудита. Это позволяет аудиторам более плодотворно проводить аудит больших объемов данных. Аудиторы могут

---

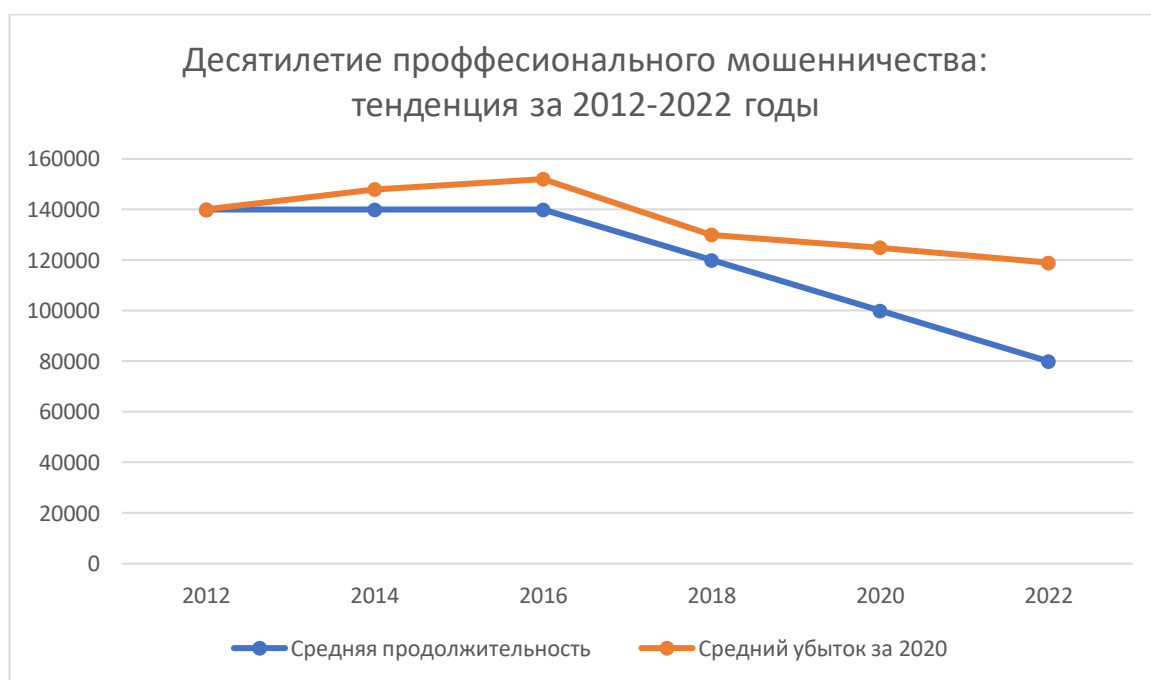
<sup>2</sup> AICPA: Audit Data Standards

извлекать данные клиентов, манипулировать ими и анализировать их. Поступая таким образом, они могут вернее познавать информацию клиента и лучше идентифицировать риски. Инструменты анализа данных позволяют изменять все данные в предварительно структурированные формы / презентации, понятные как аудиторам, так и клиентам, и даже создавать программы аудита, адаптированные к рискам четкого клиента, или передавать данные непосредственно в компьютеризированные аудиторские процедуры, что позволяет аудитору более эффективно достигать результата.

## 2. Тенденция мошенничества и коррупции.

Если говорить о важности внедрения технологического процесса в аудит, сложно не упомянуть отчет «Report to the Nations on Occupational Fraud and Abuse» (по тексту — Отчет), подготовленный Association of Certified Fraud Examiners (по тексту — ACFE) в 2022 году, в котором приводится следующая статистика:

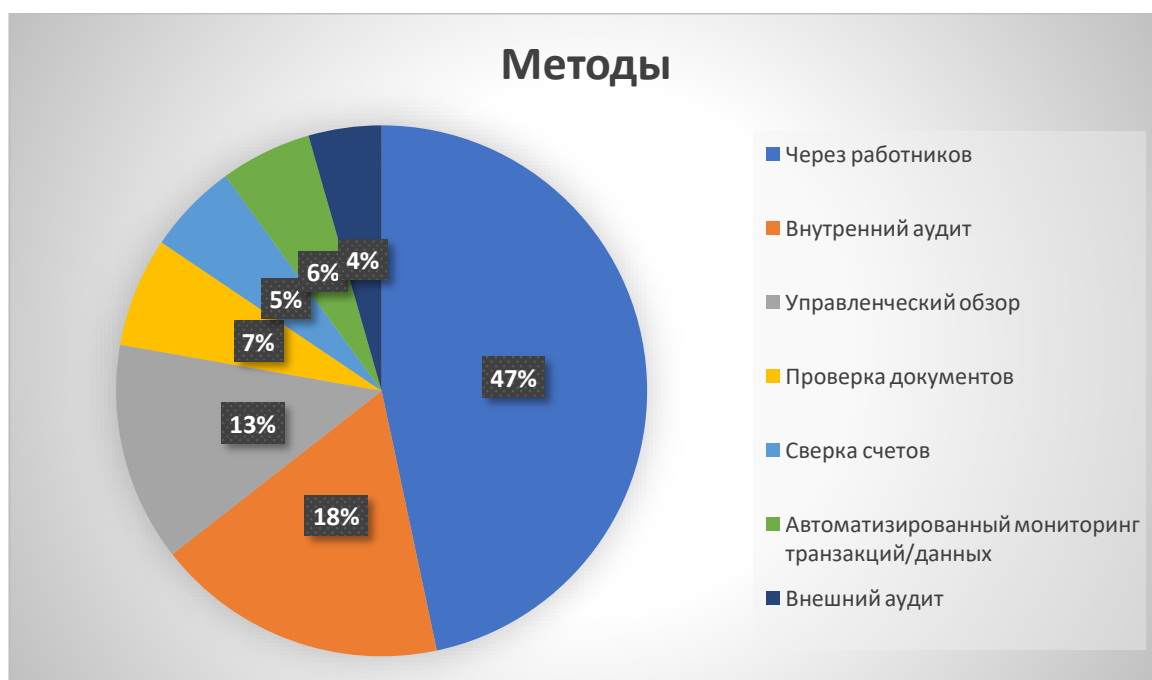
График 1 – десятилетие профессионального мошенничества: тенденция за 2012-2022 годы



По данным графика 1 мы можем сделать вывод, что мошенников стали ловить быстрее и уменьшились общие потери юридических и физических лиц. Снижение средней продолжительности на 33% и снижение потерь на 16%

Какие есть методы для первоначального выявления мошенничества?

График 2 – Методы выявления мошенничества



Данные на графике 2 подчеркивают, что при активном выявлении мошенничества оно, как правило, выявляется быстрее и приводит к меньшим потерям, а пассивное обнаружение приводит к более длительным схемам и увеличению финансового ущерба для жертвы.

Средства борьбы с мошенничеством, такие как автоматизированный мониторинг транзакций/данных, наблюдение, сверка учетных записей, постоянный и упреждающий анализ руководства и отделы внутреннего аудита, — все это инструменты, которые могут привести к более эффективному выявлению профессионального мошенничества

### 3. Крупное мошенничество Enron Corporation.

На примере Enron Corporation мы разберем как Python упрощает проверку данных.

**Enron Corporation** была американской энергетической, сырьевой и сервисной компанией, базирующейся в Хьюстоне, штат Техас. Компания была основана Кеннетом Леем в результате слияния компаний Houston Natural Gas и InterNorth в 1985 году. Она была вовлечена в такое множество сложных вещей, что люди не понимали, чем именно занимается компания. Как именно Enron зарабатывает свои деньги?

Электронная почта Enron и наборы финансовых данных — это большие беспорядочные сокровищницы информации, на которые несут огромные трудозатраты и время.

При помощи python мы сможем выполнить различные аудиторские тесты с большей эффективностью. В этом файле мы проверим обоснованность премиальных выплат и любые признаки мошенничества. Традиционно мы будем использовать метода аудита на основе выборки. В данной статье мы увидим, как точечный метод визуализации данных быстро даст нам представление о предприятии и его деятельности, а так же мы создадим идентификатор интересующих нас лиц.

Для анализа компании я использую данные 7 мая 2015 г. Версия набора данных.

Предварительный анализ данных:

Набор данных представлен в виде словаря, где ключом является имя человека, а значениями также являются словари, основанные на функциях, поскольку ключом к внутренним словарям являются функции. Следовательно, мы преобразуем словарь в dataframe. Каждый человек помечен как роі или не-роі (РОІ - Лицо, представляющее интерес)

Наш набор данных разделен на три основные категории - финансовые данные, данные электронной почты и метки роі.

Финансовые данные: «зарплата», «отсрочка платежей», «общая сумма платежей», «ссудные авансы», «бонус», «льготное приобретение акций»,

«отсроченный доход», «общая рыночная стоимость», «расходы», «прочее», «долгосрочный взнос», «ограниченные акции», «гонорары директорам» (все единицы измерения указаны в долларах США).

Данные электронной почты: «к получателю», «mail адресс», «от получателя», «от человека POI», «общий чек с POI»

Метка POI: «POI»

Общее количество точек данных: 146

Количество заинтересованных лиц: 18

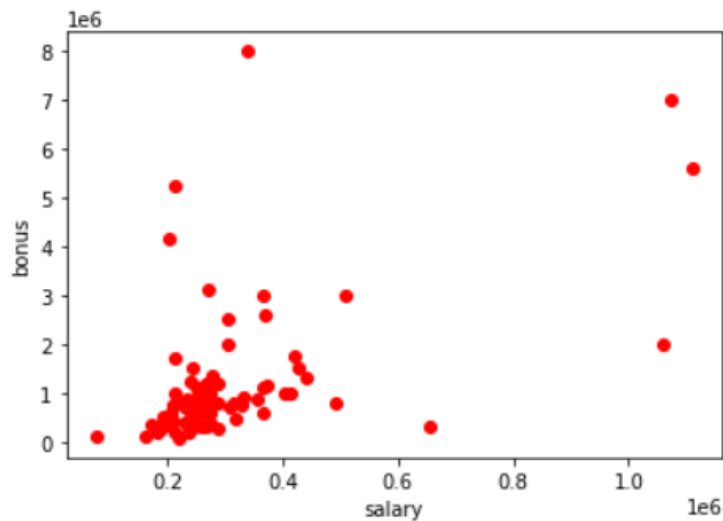
Количество лиц, не представляющих интереса: 128

Усреднение данных относится к методам для устранения нежелательного шума или поведений в данных, в то время как определение выбросов идентифицирует точки данных, которые существенно отличаются от остальной части данных.

В начале мы загружаем свои данные, чтобы программа могла их «счесть», убираем лишние данные и заполняем нулевые строки. Далее работаем уже со сформированными строками.

Теперь, когда у нас уже есть удобные для нас данные, мы легко можем их визуализировать в точечные диаграммы. Например, мы можем построить график заработной платы по x в сравнении с другим объектом по y. Сейчас мы рассмотрим сравнение заработной платы с бонусом (График – 3)

График 3 – Сравнение заработной планы с бонусами



У выбывших зарплата превышает 1000000, а бонус превышает 5000000. Теперь эти данные говорят нам, что сотрудники свалены в кучу или перемешены здесь вместе, но это обычный факт. Разберем, что тут необычного. По оси абсцисс указана заработная плата с шагом в 200000. Глядя на эти данные, я могу задать вопрос, как работники при зарплате около 400000 получает бонус в размере 8 миллионов, так же аналогично если проанализировать правую часть графика, то мы видим, что даже при максимальной зарплате свыше 1 млн они получают бонусы в гораздо большей сумме. Проведем дальнейшее расследование, чтобы ответить на вопрос о том, какова политика бонусов и почему некоторые люди получают так много.

Изучая дополнительные функции, мы получаем отклонения в столбцах «общая сумма платежей», «отложенный доход», «общий чек с ROI» и «к сообщения». Таким образом, выстраивается сюжет.

График 4 – Сравнение общих платежей и зарплаты



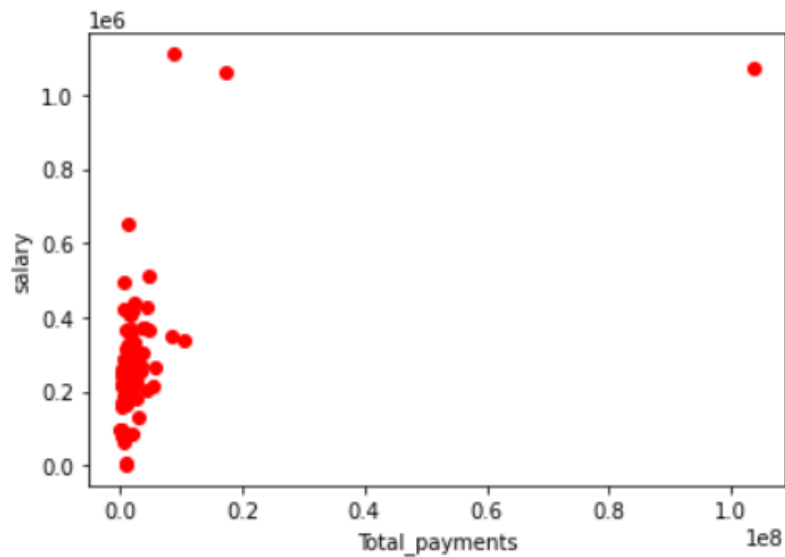
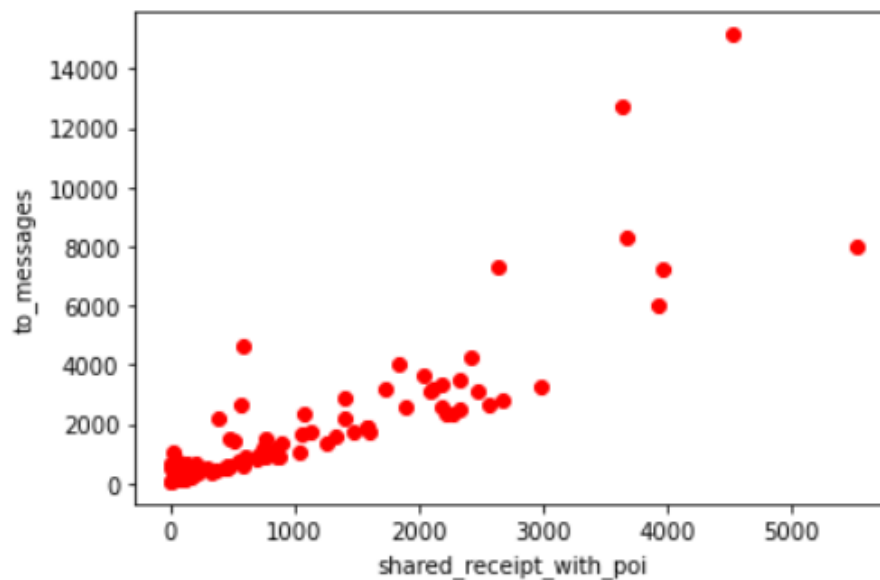


График 5 – сравнение общих чеков с ROI и отправленных сообщений



Из всех графиков сильно выделяются именно 4 точки, их мы и разберем, и выведем в виде таблицы (Таблица 1).

Таблица 1 - Подозреваемые

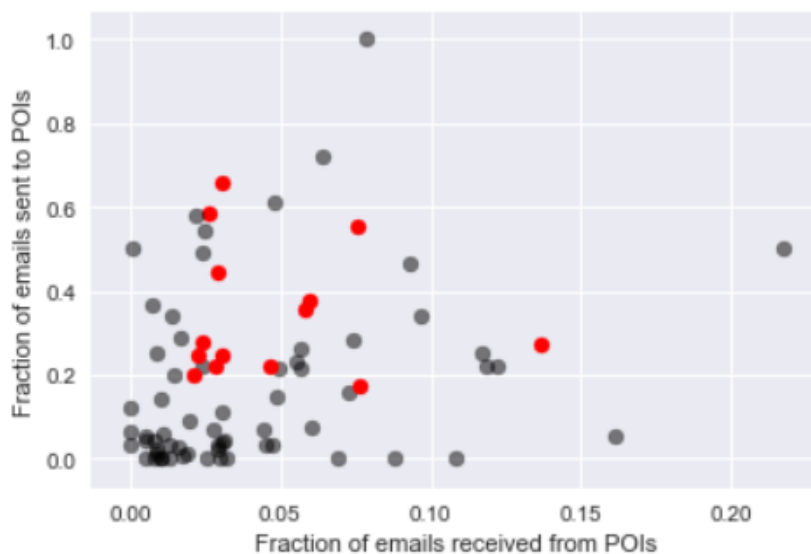
Имя	Зарплата	Бонусы	Общая сумма платежей	Общий чек с ROI	ROI
Лаворато Джон Дж.	339288	8000000	10425757	3962	Ложь

Лэй Кеннет Л.	1072321	7000000	103559793	2411	Истина
Белден Тимоти Н.	213999	5249999	550163	5521	Истина
Скиллинг Джеффри К.	1111258	5600000	8682716	2042	Истина

Следовательно, Лаворато Джон Дж., Лэй Кеннет Л., Белден Тимоти Н. и Скиллинг Джеффри К. — это 4 человека, получившие самые высокие бонусы за время работы в Enron. Что интересно, хотя Лэй и Скиллинг известны своим мошенничеством, мы получаем еще 2 человека.

Далее представим график отправленных писем ROI и полученных от ROI (График 6). Получаем такой график:

График 6– Сравнение отправленных и полученных писем от ROI



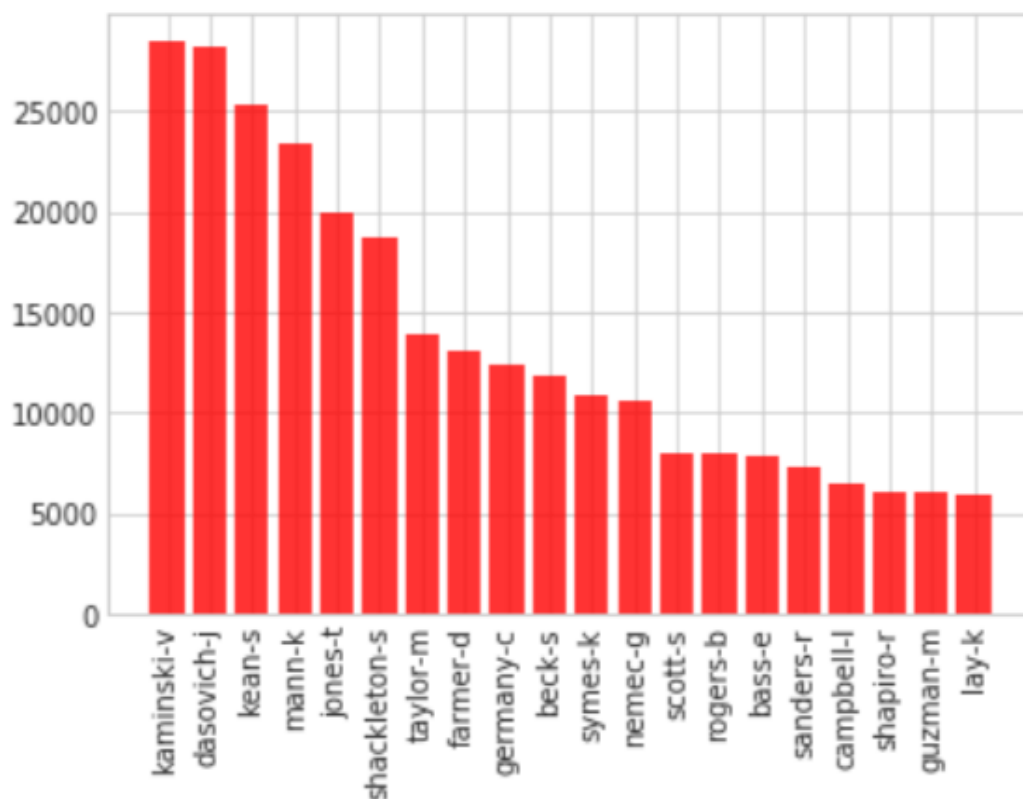
Давайте посмотрим, сколько электронных писем было отправлено каждым пользователем.

Мы составили рейтинг 20 лучших пользователей, отправивших наибольшее количество электронных писем. (График 7)

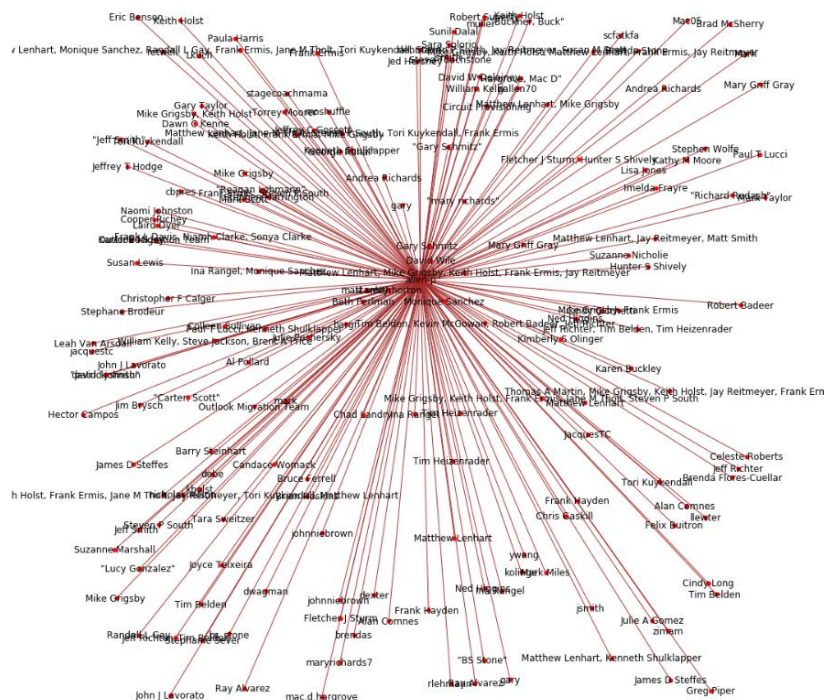
[(28465, 'kaminski-v'), (28234, 'dasovich-j'), (25351, 'kean-s'), (23381, 'man n-k'), (19950, 'jones-t'), (18687, 'shackleton-s'), (13875, 'taylor-m'), (13032, 'farmer

-d'), (12436, 'germany-c'), (11830, 'beck-s'), (10827, 'symes-k'), (10655, 'nemec-g'), (8022, 'scott-s'), (8009, 'rogers-b'), (7823, 'bass-e'), (7329, 'sanders-r'), (6490, 'campbell-l'), (6071, 'shapiro-r'), (6054, 'guzman-m'), (5937, 'lay-k)']

График 7 – 20 лучших пользователей, отправивших наибольшее количество сообщений.



Для того, чтобы больше понимать связь между количеством отправленных писем и отправителем, я хочу построить узел гистограммы (График 8), которая показывает пользователей в сравнении с количеством отправленных ими электронных писем.



На этом графике показана сеть с пользователями, которые отправляли и получали электронные письма в качестве узлов. Из этого графика можно заметить, что центральной точкой является allen-p. Следовательно, в наборе из 1000 электронных писем allen-p кажется влиятельным.

Далее, чтобы правильно предположить, надо ли подозревать человека, исходя из любых признаков, характеризующих его работу, мы будем пользоваться классификацией.

Мы собираемся использовать два оценочных показателя: точность и отзыв. В контексте этого проекта:

Точность измеряет, сколько положительных прогнозов было на самом деле положительными наблюдениями, то есть доля предсказанных ROI, которые на самом деле были неверны.

Отзыв измеряет, сколько положительных наблюдений было предсказано правильно, то есть доля фактических ROI, которые были предсказаны правильно.

В итоге после недолгих махинаций с кодом мы выявляем людей, который попадут в список первых проверяющих.

В ходе судебного разбирательства этого дела в 2001 году 18 человек (в том числе Джефф Скиллинг и Кеннет Лей) были осуждены.

Этот анализ был сосредоточен на финансовых данных и для примера взяты всего 1000 писем, но в нашем распоряжении имеется огромный массив электронных писем (более 500000), а также при дальнейшем расследовании мы можем посмотреть журнал посещаемости сверхурочных сотрудников-призраков и часы работы каждого персонала. Анализ этой части данных с целью выявления закономерностей, помог бы и, вероятно, дал бы нам больше материала для обнаружения и правильной маркировки мошенников.

Благодаря этому примеру мы смогли наглядно посмотреть, как Python используется в аудиторской деятельности, как упрощает и ускоряет работу, а также способствуем постепенному вводу информационных технологий во все сферы человеческой жизни!

## Литература

1. Федеральный закон от 01.07.2010 N 136-ФЗ, от 01.12.2014 N 403-ФЗ, от 02.07.2021 N 359-ФЗ  
Режим доступа: [http://www.consultant.ru/document/cons\\_doc\\_LAW\\_83311/5662200f71fec9b8ab1a6e0615896613d3852b5a/](http://www.consultant.ru/document/cons_doc_LAW_83311/5662200f71fec9b8ab1a6e0615896613d3852b5a/) (Дата обращения: 15.01.2023)
2. Bethany McLean «The Smartest Guys in the Room: The Amazing Rise and Scandalous Fall of Enron»  
Режим доступа: <https://www.labyrinth.ru/books/886742/> (Дата обращения 15.01.2023)
3. ORD Occupational Fraud 2022: A Report to the Nations. Режим доступа: <https://acfepublic.s3.us-west-2.amazonaws.com/2022+Report+to+the+Nations.pdf> (Дата обращения: 15.01.2023)
4. AICPA Audit data standard: Режим доступа: <https://www.aicpa.org/resources/download/base-standard-audit-data-standards> (Дата обращения 15.01.2023)
5. Интернет ресурс: <https://github.com/MahnoorJaved98/Enron-Case> (Дата обращения: 15.01.2023)
6. Интернет ресурс: <https://github.com/AICPA-AuditDataAnalytics2018/ADS---Python-Example-> (Дата обращения: 15.01.2023)
7. Интернет ресурс: <https://www.kaggle.com/code/ankur561999/data-cleaning-enron-email-dataset> (Дата обращения: 15.01.2023)

## Literature

1. Federal Law No. 136-FZ of 01.07.2010, No. 403-FZ of 01.12.2014, No. 359-FZ of 02.07.2021. Access mode: [http://www.consultant.ru/document/cons\\_doc\\_LAW\\_83311/5662200f71fec9b8ab1a6e0615896613d3852b5a/](http://www.consultant.ru/document/cons_doc_LAW_83311/5662200f71fec9b8ab1a6e0615896613d3852b5a/) (Accessed: 01/15/2023)

2. Bethany McLean "The Smartest Guys in the Room: The Amazing Rise and Scandalous Fall of Enron» Access mode: <https://www.labyrinth.ru/books/886742/> (Accessed 15.01.2023)

3. ORD Occupational Fraud 2022: A Report to the Nations. Access mode: <https://acfepublic.s3.us-west-2.amazonaws.com/2022+Report+to+the+Nations.pdf> (Accessed: 15.01.2023)

4. AICPA Audit data standard Режим доступа: <https://www.aicpa.org/resources/download/base-standard-audit-data-standards> (Дата обращения 15.01.2023)

5. Internet resource: <https://github.com/MahnoorJaved98/Enron-Case> (Accessed: 15.01.2023)

6. Internet resource: <https://github.com/AICPA-AuditDataAnalytics2018/ADS---Python-Example> - (Accessed: 15.01.2023)

7. Internet resource: <https://www.kaggle.com/code/ankur561999/data-cleaning-enron-email-dataset> (Accessed: 15.01.2023)