

Министерство образования и науки Российской Федерации

Университет «Дубна»

Институт системного анализа и управления

Прикладная информатика

Практическая работа №3

Big Data

Выполнил студент

Группы №1074

Семина В.В.

Проверил преподаватель

Куликов Д.Л.

г. Дубна 2020

# Задание 1

**Преобразование 4-х буквенных слов при помощи генетического алгоритма**

Начальное слово* <b>лупа</b>	Конечное слово* <b>ромб</b>	Максимальная длина цепочки <b>200</b>
---------------------------------	--------------------------------	--

Слово из которого надо получить второе слово

Слово в которое требуется преобразовать первое слово

Размер популяции  
**150**

Максимальное количество элементов в каждом поколении

**РАСЧИТАТЬ**

**Результат**

Слово	Значение	Шаг
ЛУПА	Собирающая линза или система линз с небольшим фокусным расстоянием (10-100 мм).	1
ЛУНА	1) Богиня ночного света в древнеримской мифологии. 2) Естественный спутник Земли, находится от нее на среднем расстоянии 384 400 км.	2
КУНА	1) Весовая и денежная единица, а также название монет, использовавшихся на Руси в X–XV веках. 2) Денежная единица Хорватии.	3
КУМА	1) Крестная мать по отношению к родителям крестника и к крестному отцу. 2) Река на Северном Кавказе. 3) Река в Западной Сибири, правый приток реки Конда.	4
КУМБ	Небольшое финикийское рыболовное судно.	5
РУМБ	Направление к точкам видимого горизонта относительно стран света или угол между двумя такими направлениями.	6
РОМБ	Геометрическая фигура, параллелограмм с равными сторонами.	7

Размер страницы: **20** 1 - 7 < >

**Преобразование 4-х буквенных слов при помощи генетического алгоритма**

Начальное слово* <b>лупа</b>	Конечное слово* <b>ромб</b>	Максимальная длина цепочки <b>10</b>
---------------------------------	--------------------------------	---

Слово из которого надо получить второе слово

Слово в которое требуется преобразовать первое слово

Размер популяции  
**500**

Максимальное количество элементов в каждом поколении

**РАСЧИТАТЬ**

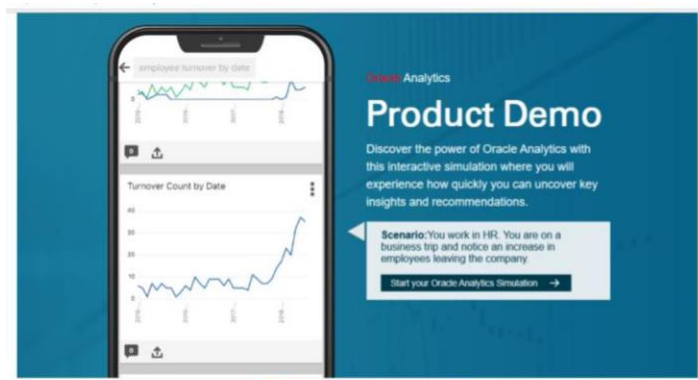
**Результат**

Слово	Значение	Шаг
ЛУПА	Собирающая линза или система линз с небольшим фокусным расстоянием (10-100 мм).	1
ЛУНА	1) Богиня ночного света в древнеримской мифологии. 2) Естественный спутник Земли, находится от нее на среднем расстоянии 384 400 км.	2
КУНА	1) Весовая и денежная единица, а также название монет, использовавшихся на Руси в X–XV веках. 2) Денежная единица Хорватии.	3
КУМА	1) Крестная мать по отношению к родителям крестника и к крестному отцу. 2) Река на Северном Кавказе. 3) Река в Западной Сибири, правый приток реки Конда.	4
КУМБ	Небольшое финикийское рыболовное судно.	5
РУМБ	Направление к точкам видимого горизонта относительно стран света или угол между двумя такими направлениями.	6
РОМБ	Геометрическая фигура, параллелограмм с равными сторонами.	7

Размер страницы: **20** 1 - 7 < >

Для преобразования слова «ЛУПА» в слово «РОМБ» с применением генетического алгоритма необходимо и достаточно использовать длину цепочки равную 7 и размер популяции – 35. При размере популяции 31 можно получить генетический алгоритм длиной цепочкой в 9 шагов. Но при увеличении популяции до 35 алгоритм сокращается на 2 шага. При увеличении популяции количество шагов не изменится. Следовательно, достаточный размер популяции – 35.

## Задание 2

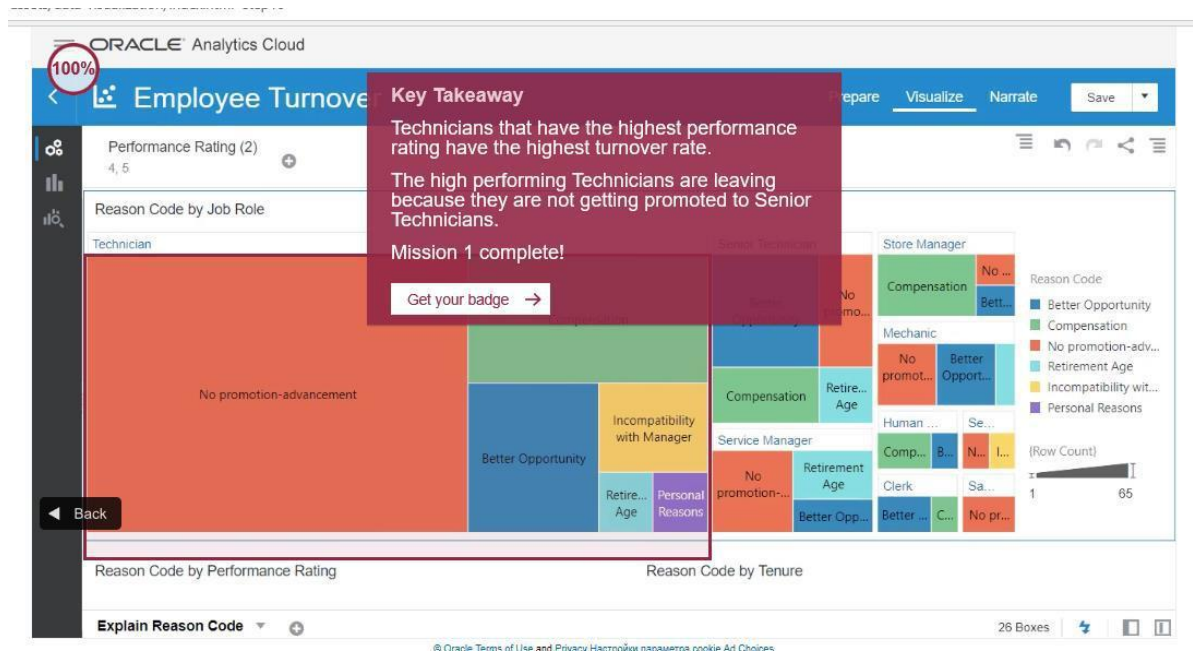


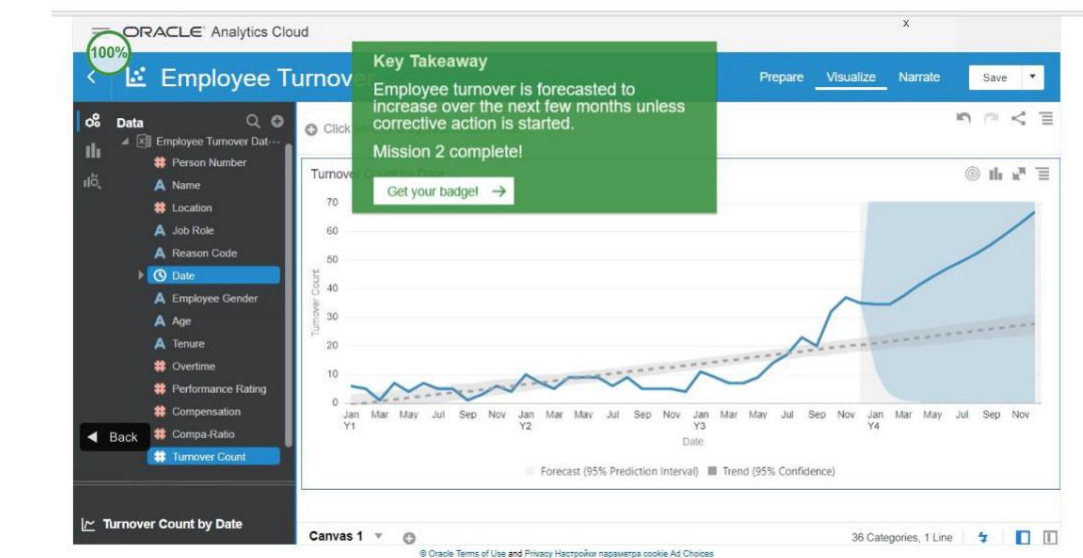
**Your first mission awaits!**

**1**

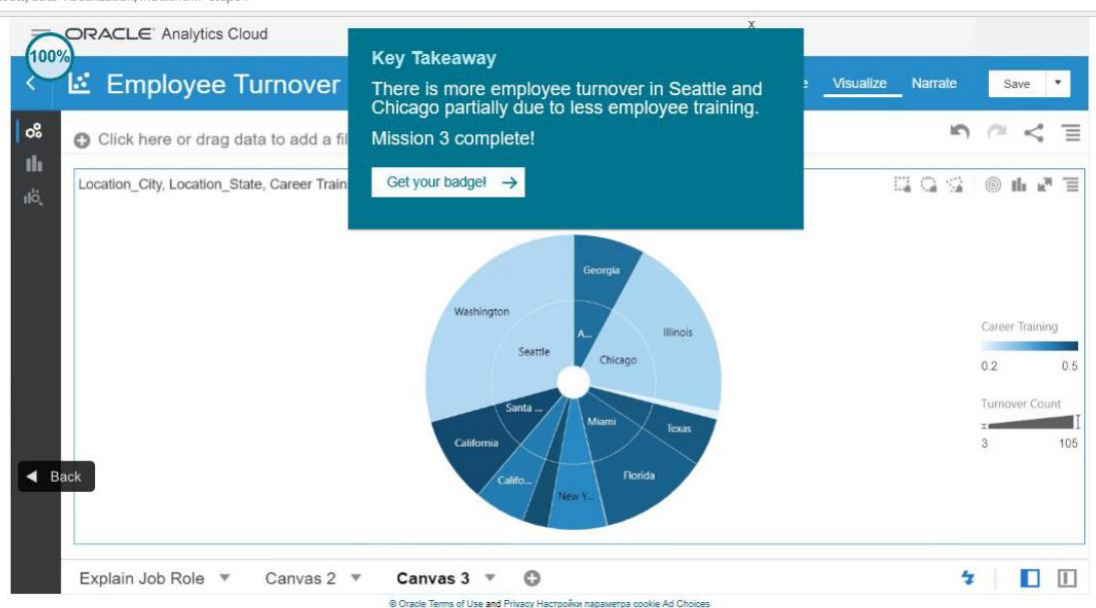
- Mission 1:** Quickly understand why employees are leaving your company (2 weeks)  
**Purpose:** Discover how Oracle Analytics quickly identifies key insights with employee turnover data.
- Mission 2:** Understand and forecast the future turnover rate (2 weeks)  
**Purpose:** Quickly forecast employee turnover rate for the next few quarters.
- Mission 3:** Discover additional insights to address and reverse the employee turnover trend (2 weeks)  
**Purpose:** Add new data identifying which employees have completed employee training and where.

[Begin mission 1](#)





/assets/data-visualization/index.html#step34



Зарегистрировавшись на сайте, самой главной задачей нужно было выявить причины высокой текучести кадров. Сделав анализ на первом этапе можно убедиться, что одной из главных причин высокой текучести кадров является отсутствие продвижения на службе. Это стало основной причиной увольнения 50% рабочих. Так же в причинах было указано недовольство заработной платой.

На втором этапе работы нужно было рассмотреть и проанализировать текучесть кадров. На графике показано количество рабочих, ушедших по собственному желанию. Печальным фактом является высокий темп развития увольнений и значительное превышения линии тренда, что наталкивает на мысль о срочном изменении системы.

В третьем этапе был выведен график, показатели которого говорят о текучести кадров городах из-за отсутствия стандартных навыков работников. Таким образом, можно сказать, что в Сиэтле и в Чикаго сотрудники уволены из-за меньшей подготовки.

### Задание 3

Дашборд – одна из возможностей платформ Business Intelligence. Это крупные системы, которые переводят данные бизнеса в доступную информацию. Визуальное представление такой информации и называется дашбордом.

Tableau специализируется на анализе данных через их визуализацию для решения вопросов Business Intelligence.

В нем легко создавать интерактивные дашборды, которые позволяют изучить динамику, тренды и структуру данных, с помощью графиков.

Ключевое отличие — комбинировании данных из разных БД и источников. Также платформа позволяет нескольким пользователям одновременно работать над отчетом в реальном времени.

Tableau обладает широчайшими возможностями визуализации: богатая библиотека платформы включает в себя облака слов, пузырьковые и древовидные диаграммы, которые позволяют достигнуть более высокого уровня понимания своих данных и их контекста.

## Задания 4

### Рекомендательный движок

E-commerce/ритейл/маркетинг



Пример: На основе выбора покупателя предлагать сопутствующие товары «с этим товаром покупают»

С помощью программы 1C и PostgreSQL производится анализ и собирается информация о том, чего хочет покупатель. А именно акцент делается на товары, которые рассматривал покупатель. Вследствие чего делаем вывод, что он заинтересован в них.

PostgreSQL – объектно-реляционная система управления базами данных. То есть вся собранная информация распределяется и хранится в базах данных.

Далее все данные передаются в Apache Spark. Это фреймворк для реализации распределённой обработки неструктурированной и слабоструктурированных данных. Одним словом на данном этапе вся информация собирается, распределяется и структурируется.

Дальше все переходит в Hotbox, который рассчитан на хранение и быструю раздачу большого количества данных. На этом этапе идет распределение, запись и хранение большого количества информации данных для дальнейшего использования.

Следующее движение в этой цепочке происходит в Spark и TensorFlow. Это все открытые библиотеки для машинного обучения. Цель этих программ автоматически находить и классифицировать образы продуктов, на которые обратил внимание покупатель, и подбирать похожие материалы.

Далее через оставшиеся ресурсы все попадает к человеку в рекомендация.