

Домашняя работа 1

Вариант 1. Кредитные рейтинги

Файлы содержат данные 1) о присвоении рейтинговыми агентствами рейтингов российским компаниям (рейтинговых действиях); 2) о кредитных событиях, связанных с компаниями; 3) о числовой кодификации одной из рейтинговых шкал.

1. Запросами SQL загрузить¹ в БД данные из файлов **ratings_task.xlsx**, **credit_events_task.xls** и **scale_EXP_task.csv**. Выбрать форматы полей, создать таблицы (наложить разумные ограничения на поля), импортировать данные. (1 балл)
2. Запросами SQL вынести из таблицы **actions** информацию о рейтингах (тип, срочность и т.д.) и информацию о рейтингуемом лице (текстовое название и лишние числовые коды) в отдельные таблицы. (1 балл)
3. Связать таблицы в базе, задав внешние ключи, где это целесообразно. (1 балл)
4. В сформированной базе данных составьте запрос, который выводит для выбранных вида рейтинга (**rat_id**) и даты (**date**) в таблице **actions**, все актуальные рейтинги. Актуальным рейтингом признается значение, присвоенное или подтвержденное последним рейтинговым действием до исследуемой даты, при условии, что это действие было не снятием и не приостановкой рейтинга. Таблица должна содержать поля **ent_name**, **grade** и **assign_date** (дату присвоения актуального рейтинга). (2 балла)

Данные: см. ссылку в письме на групповой почте.

Вариант 2. Облигации

Файлы содержат данные о 1) спецификациях облигаций российского рынка; 2) о котировках облигаций из разных источников; 3) о листинге бумаг на разных платформах.

1. Запросами SQL загрузить² в БД данные из файлов **bond_description_task.xlsx**, **quotes_task.xlsx** и **listing_task.xlsx**. Выбрать форматы полей, создать таблицы (наложить разумные ограничения), импортировать данные. (1 балл)
2. Запросами SQL вынести из таблицы **bond_description** информацию об эмитентах и из таблицы **quotes** информацию о площадке (и поместить ее в таблицу **listing**). (1 балл)
3. Связать таблицы в базе, задав внешние ключи, где это целесообразно. (1 балл)
4. В сформированной базе данных составьте запрос, возвращающий имена эмитентов, чьи облигации торгуются на МБ в режиме основных торгов, и по каждой из которых доля дней, когда котировка [BID или ASK выберете самостоятельно] не существовала, составляла бы не более 10% наблюдений за бумагой. Таблица должна содержать поля **issuer**, **ISIN** и **nun_ratio** (долю наблюдений без пропуска котировки). (2 балла)

Данные: см. ссылку в письме на групповой почте.

¹ Перед загрузкой в БД файлы можно модифицировать, не нарушая структуру содержащихся данных, т.е. можно изменять форматы значений, изменять типы файлов (расширения) и т.д., но не разделять таблицы, не удалять/добавлять/разделять столбцы и т.д.

² Перед загрузкой в БД файлы можно модифицировать, не нарушая структуру содержащихся данных, т.е. можно изменять форматы значений, изменять типы файлов (расширения) и т.д., но не разделять таблицы, не удалять/добавлять/разделять столбцы и т.д.

Материалы для сдачи

1. Краткая инструкция, содержащая:
 - а. подготовительные действия для работы с данными (где и в каком виде хранить данные, какой софт установить и т.д.);
 - б. краткое содержания процедур, заложенных в скриптах/запросах, а также порядок их запуска;
 - с. описание ожидаемых результатов каждого этапа.
2. Скрипты с запросами в формате sql.
3. Все (кроме данных) необходимо разместить в репозиторий на github, в котором преподаватель является участником с правом редактирования.

Требования к выполнению задания

1. Все процедуры и результаты должны быть воспроизводимы и описаны в мере, обеспечивающей использование третьими лицами.

Правила и сроки выполнения

1. Домашняя работа является индивидуальной активностью студента.
2. Срок выполнения (первая итерация сдачи) Домашней работы 1: **17 ноября 23:59:59**.
3. Максимальная оценка за выполненное задание: **5 баллов**.
4. По результатам первой итерации студент получает фидбэк и предварительную оценку. Ориентировочно через **3 дня после сдачи работы**. Получив фидбэк, студент может исправить работу и сдать ее до **19 декабря**, но не позднее, чем другие домашние работы по курсу.
5. Студент пропустивший дедлайн первой итерации сдает работу на итоговую проверку **без возможности ее исправить**.
6. Студент выполнивший все пункты Домашней работы 1 (не зависимо от итерации) получает право проверки аналогичной работы сокурсника. Срок проведения peer-review: **4 дня с момента получения задания на проверку**. Продление сроков peer-review допускается по договоренности с преподавателем.
7. Peer-review проводится по следующим аспектам:
 - *Полнота* (по всем ли пунктам была предпринята попытка выполнения?)
 - *Понятность* (инструкция понятно написано? понятно ли прописаны требования к исходным данным? понятно ли, что делает код и каким результатам это приводит? пользовательские параметры обособлены и понятны?)
 - *Содержание* (сделано именно то, что требуется в задании? если нет, то в чем отличия того, что сделано? в чем это выражается и что Вы сделали, чтобы это увидеть?)
 - *Корректность* (разумно ли выбраны форматы поле таблиц? присутствуют ли ошибки исполнения кода или смысловые ошибки? в чем именно ошибка? как Вы это обнаружили? знаете ли Вы как это исправить? видите ли Вы заведомо абсурдные результаты?)
8. Максимальная оценка за peer-review: **4 балла**.