Добрый день, хотела бы представить проект на тему «Прогнозирование объемов закачанного и выкачанного газа».

СЛАЙД

Начнем с поставленной задачи. В данном случае необходимо было построить модель, которая позволяла бы получать прогноз объемов закачанного и выкачанного газа для каждой страны на неделю вперед.

СЛАЙД

Работа началась с ознакомления с данными. У меня имелась информация о соответствии ПХГ определенной стране. Уточню, что аббревиатура ПХГ — Подземное хранилище газа.

Чтобы при последующей обработке данных без особых усилий можно было заменить название ПХГ на название страны, я делаю словарь, в котором ключ это id ПХГ, а значение - страна.

СЛАЙД

Далее перейдем к основному блоку данных. У меня имелась информация по объему закаченного газа, которые выглядели следующим образом. Первая колонка соответствует дате, далее следует 120 колонок с префиксом VGS, которые соотвествуют разным подземным хранилищам газа.

ЩЕЛЧОК

Здесь работаю с пропусками в данных. Удаляю полностью пустые столбцы. Оставшиеся пустые значения заменяем на ноль.

ЩЕЛЧОК

После чего используем словарь соответствия ПХГ и стран, созданный ранее, чтобы изменить названия колонок и объединяем столбцы, принадлежащие одной стране, суммируя их.

СЛАЙД

В итоге получается такой результат. Все эти операции объединяем в функции, которые применяю так же к данным по выкаченному газу.

СЛАЙД

Далее создается таблицу баланса ПХГ, вычитая из данных закаченного данные выкаченного газа.СЛАЙД

На баланс ПХГ могут оказывать влияние разные факторы. К которым также могут относится погодные условия. У меня была таблица тикеров. Тикер – код чего-либо, в данном случае это год города. Тогда тикер соответствует какому-то городу, а следовательно и стране. Я так же сделала словарь соответствия, в котором ключ это тикер, а значение – страна.

СЛАЙД

Среди природных факторов у меня имелись данные по среднесуточной температуре воздуха в городах, в которых природный газ используется в промышленности, по количеству выпавших осадков, по скорости ветра. Файл экселя был оформлен следующим образом, поэтому для начала необходимо было преобразовать данные к удобному для нас виду.

После чего я убрала пустые значения, изменила тикеры при помощи словаря на название страны и объединила столбцы с одинаковым названием, взяв среднее. Все эти действия я провела для всех природных факторов. То есть помимо температуры были обработаны данные по скорости ветра и осадкам.

СЛАЙД

На этом начальная обработка данных была завершена и я решила посмотреть как сильно влияют перечисленные ранее погодные факторы на баланс ПХГ.

Была отражена матрица попарной корреляции. По которой можно было заметить, что средняя температура имеет тесную линейную связь с объёмами газа. Что касалось ветра и осадков, то тут было не всё так радужно. Было выдвинуто предположение, что данные природные факторы оказывают влияние с каким-то лагом.

СЛАЙД

Для этого была написана функция по поиску наилучшего лага для каждой страны. После чего был создан словарь, в котором каждой стране сопоставлялся лучший лаг по погоде и по осадкам.

СЛАЙД

Далее непосредственно перейдем к моделированию предсказаний. Дальнейший рассказ буду вести на примере Австрии, хотя аналогичная ситуация была бы и у других стран.

Для моделирования и предсказания я использовала Профит. Смотрела я на четыре модели: предсказание баланса ПХГ, предсказание баланса при экзогенной температуре, предсказание баланса при трех экзогенных факторах, предсказание при трех экзогенных факторах и лучших лагах. Эти же варианты были проделаны, но не сразу для баланса, а отдельно для выкаченного и вкаченного газа, а после этого уже рассматривалась их разница Качество прогнозирования я оценивала при помощи МSE, которые соответственно были равны значениям, представленным на слайде. Можно сказать, что при прогнозировании баланса пхг результат получался хуже, чем при прогнозировании отдельно объемов вкаченного и выкаченного газа, а потом вычисления разности. А наилучший вариант получался при добавлении трех экзогенных факторов в модель без лучших лагов.

СЛАЙД

Однако, ошибка казалась очень большой. поэтому я решила продолжить обработку данных и поработать с выбросами. Если до этого данные по Австрии имели следующий вид, то после того как я заменила значения выше 95 и ниже 5% квантили на полу сумму соседних значений, я уже получила следующий результат.

СЛАЙД

Ошибки моделей тут же улучшились и приняли следующие значения.

Также в качестве эксперимента была добавлена недельная сезонность, которая дала рекордное уменьшение ошибки.

В итоге удалось уменьшить с 24 тысяч до 14.

СЛАЙД

Все эти модели занимали достаточное количество места и файл получался очень большим, причем эти модели запускались только для определенной страны. Чтобы оптимизировать запуск моделей, мне посоветовали для удобства оформить все как объектно-ориентированное программирование.

Я сделала класс VGS\_OF\_COUNTRY. У экземпляров этого класса доступны следующие методы. За исключение первого, который является статическим.

При создании экземпляра класса идет проверка на то, существуют ли данные для выбранной страны, если таких данных не существует, то будет выдаваться ошибка.

СЛАЙД

Таким образом, если раньше приходилось все запускать через множество функций, то теперь у нас есть метод класса, при помощи которых я сразу же могу отразить ошибку модели, изобразить прогноз, или же сравнить полученный прогноз с исходными значениями. И все это можно делать меняя параметры у метода.

СЛАЙД

Так же была добавлена возможность создания прогноза на произвольный период времени и с желаемой даты. Однако есть ограничение при создании модели, дата прогноза не может быть позже 9 марта 2020 года, соответственно и период прогноза не заходить за эту дату.

Переходные времена года осень и весна прогнозируются хуже, чем лето и зима. На данном слайде представлено сравнение исходных данных и прогноза на одну летнюю неделю для Германии, Франции, Великобритании, Нидерландов, Италии, Австрии и Венгрии.

СЛАЙД

Для этих же стран прогноз на 2 недели в зимний период выглядит следующим образом.

СЛАЙД

В переходных временах года могут быть неточности в прогнозе для определенной страны. Я добавила возможность выбрать баланс ПХГ не только отдельно выбранной страны, но и для всех хранилищ подземного газа в Европе. В таком случае прогноз для осеннего периода получается лучше, чем в каждой стране по отдельности.

СЛАЙД Это пожалуй все, что я хотела рассказать, спасибо за внимание!