

Exploratory data analysis

312 группа
Евенко, Ефарова, Коротчук

Москва, 2019

Формулировка задачи

Необходимо решить вопрос о том, с какой из компаний следует заключить эксклюзивный договор на поставку стали.

Имеется два основных поставщика стали - Westeros Inc. и Harpy & Co.

Дан CSV-файл с данными о производстве оружия и количестве единиц сломанного оружия за каждый месяц каждым из кузнецов.

- Анализ csv-файла
- Выбор техники исследования
- Написание программы с демонстрацией выбранной техники
- Выводы на основании полученных результатов

Вид исходной таблицы:

	A	B	C	D	E	F
1	<u>unsullien.id</u>	<u>production_date</u>	<u>report_date</u>	<u>produced</u>	<u>defects</u>	<u>supplier</u>
2	1.0	1	1	103.0	0.0	harpy.co
3	1.0	1	2	0.0	2.0	harpy.co
4	1.0	1	3	0.0	4.0	harpy.co
5	1.0	1	4	0.0	5.0	harpy.co
6	1.0	1	5	0.0	13.0	harpy.co
7	1.0	1	6	0.0	11.0	harpy.co
8	1.0	1	7	0.0	11.0	harpy.co
9	1.0	2	2	105.0	0.0	harpy.co
10	1.0	2	3	0.0	2.0	harpy.co

- A- ID кузнеца
- B- дата производства партии мечей, о которой идёт речь в текущем отчете
- C- дата отчета
- D- количество произведенных мечей в день отчета
- E- количество испорченных мечей в день отчета
- F- поставщик стали

Exploratory data analysis

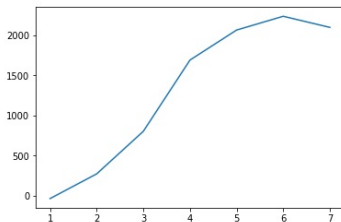
В статистике разведывательный анализ данных (РАД) - это подход к анализу наборов данных для обобщения их основных характеристик, зачастую с помощью визуальных методов. Статистическая модель может использоваться или нет, но в первую очередь EDA предназначена для того, чтобы увидеть, что данные могут показать нам, помимо формальных задач моделирования или проверки гипотез.

Одной из техник РАД является график прогона (Run chart), который мы используем в нашей работе. Он отображает данные наблюдений во временной последовательности.

Полученные результаты

- Построен график, показывающий разность между оставшимися мечами из стали Harpy & Co и оставшимися мечами Westeros Inc. за временной период равный 7 месяцам
- Построены графики, показывающие объемы производства каждой компании в каждом месяце
- Анализ каждой партии производства. Для партии первого месяца строятся графики, показывающие изменение отношения суммарного количества поломок за текущий месяц и предыдущие к объему партии. Также для всех партий построены графики, показывающие изменение отношения количества поломок за текущий месяц к объему партии

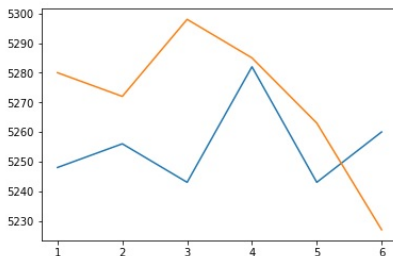
Разность между оставшимися мечами



Делаем вывод о том, что количество оставшихся мечей из стали компании Harpy & Co больше, чем количество оставшихся мечей Westeros Inc., так как график положительный. Но в последнем месяце график начинает убывать, что позволяет сделать предположение о том, что при долгосрочном договоре первая компания будет все же менее выгодна.

Объемы производства

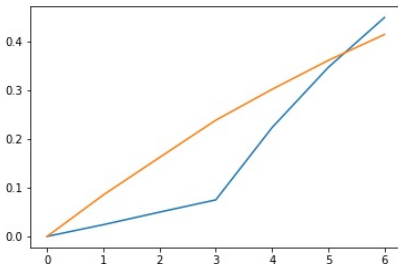
Синяя кривая - Harpy & Co
оранжевая - Westeros Inc.



На основании этих данных мы можем сделать вывод о том, что количество производимых мечей в каждом месяце примерно совпадает. То есть следующим шагом можно анализировать отношение объема поломок к объему производства.

Анализ партий производства

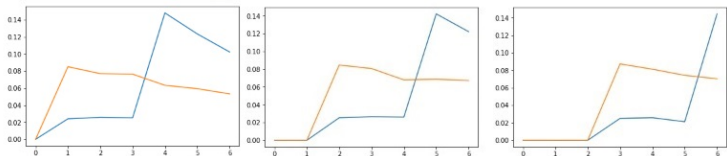
Синяя кривая - Harpy & Co
оранжевая - Westeros Inc.



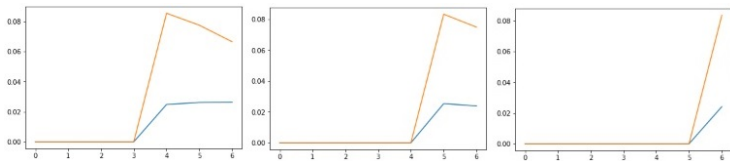
Изменение отношения суммарного количества поломок первой партии за текущий месяц и предыдущие к объему первой партии.

Партия мечей из стали Harpy & Co сначала несколько месяцев имеет медленно растущий график, но через несколько месяцев мы наблюдаем резкое возрастание этой величины. Причем к последнему месяцу эта величина у компании Harpy & Co не только продолжает также возрастать, но и становится больше этой же величины у компании Westeros Inc.

Анализ партий производства



Первая, вторая и третья партии



Четвертая, пятая и шестая партии

Так как выпуски мечей производились в первые 6 месяцев, то имеем 6 партий оружия. На каждом рисунке мы видим подтверждение нашему предположению. Каждая партия мечей из стали Harpy & Co сначала несколько месяцев имеет медленно растущее отношение количества поломок за текущий месяц к объему этой партии, но через несколько месяцев мы наблюдаем резкое возрастание этой величины. Причем эта величина у компании Harpy & Co не только резко возрастает, но и становится больше этой же величины у компании Westeros Inc.. На основе данных графиков делается вывод о схожем поведении каждой партии обеих компаний.

Продemonстрированные графики подтверждают предположение о том, что следует заключить договор с компанией Westeros Inc.. Так как несмотря на хорошие показатели каждой партии стали Harpy & Co в первые месяцы после их производства, мы замечаем их тенденцию к значительным поломкам спустя несколько месяцев. И основываясь на наших прогнозах в дальнейшем мы будем иметь больше исправного оружия, заключив сделку с Westeros Inc..