|  |  |  |
| --- | --- | --- |
| Машинное обучение и анализ данных | | |
| Кафедра САПРиПК  Лабораторная работа №4  «Метод главных компонент» | ФИО студента | Козякин А.Д., Орлов Н.Е., Полешко Р.А. |
| Группа | ИВТ-363 |
| Дата выполнения |  |
| Дата отчета |  |
| ФИО преподавателя | Фокин Р.О. |
| Оценка |  |
| Подпись преподавателя |  |

Загрузка и предобработка данных(используется датасет про швейную промышленность).

**Информация об атрибутах**:

1) date: Дата в ММ-ДД-ГГГГ.

2) day: День недели.

3) quarter: Часть месяца. Месяц был разделен на четыре квартала.

4) department: Ассоциированный отдел с экземпляром.

5) team: Связанный номер команды с экземпляром

6) no\_of\_workers: Количество работников в каждой команде

7) no\_of\_style\_change: Количество изменений в стиле конкретного продукта

8) targeted\_productivity: Целевая производительность, установленная Органом для каждой команды на каждый день.

9) smv: Стандартное значение минуты, это выделенное время для задачи

10) wip: Незавершенное производство. Включает количество незавершенных позиций для продуктов

11) over\_time: Представляет собой количество сверхурочной работы каждой команды в минутах

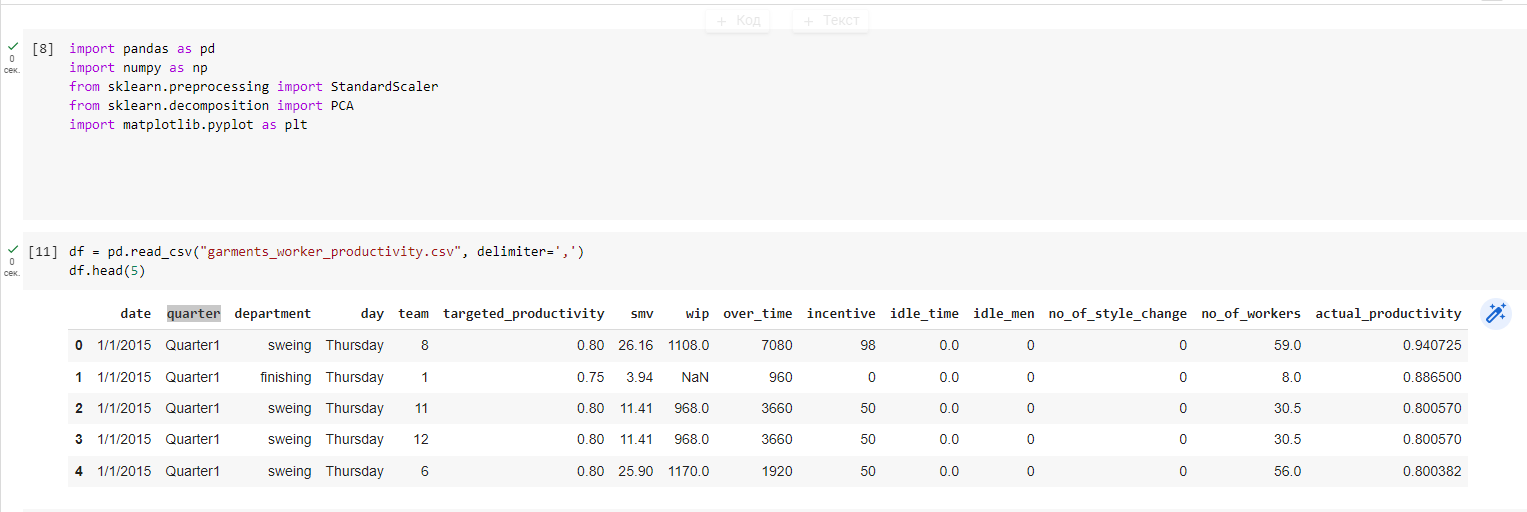
12) incentive: Представляет собой сумму финансового стимула (в БРЭ), который позволяет или мотивирует определенный курс действий.

13) idle\_time: Количество времени, когда производство было прервано по нескольким причинам

14) idle\_men: Количество рабочих, которые простаивали из-за перерыва

в производстве

15) actual\_productivity: фактический процент производительности, которую обеспечивали рабочие. Он колеблется от 0-1.

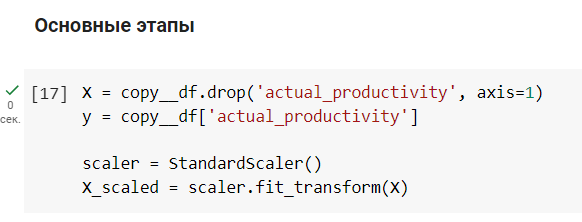


Удаление лишних полей:

Изображение выглядит как стол

Автоматически созданное описание

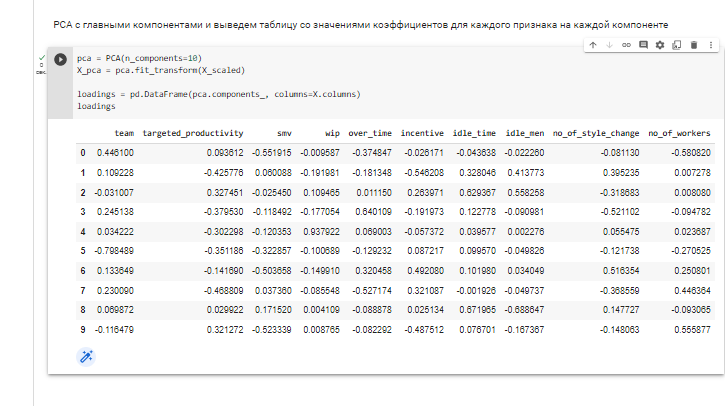
Далее, применим метод главных компонент (PCA) и построим график кумулятивной доли, объясненной дисперсии для определения количества главных компонент, которые необходимо оставить:



Изображение выглядит как диаграмма

Автоматически созданное описание

Далее, применим PCA с главными компонентами и выведем таблицу со значениями коэффициентов для каждого признака на каждой компоненте:



Для проведения анализа главных компонент на выбранных параметрах, выберем параметры over\_time (количество сверхурочной работы каждой команды в минутах) и idle\_time (количество времени, когда производство было прервано по нескольким причинам):

Изображение выглядит как диаграмма

Автоматически созданное описание