

## Вопросы:

### 1) Рынок DE: Количество вакансий, отрасли компаний, размер компаний:

На сайте hh.ru порядка 960 вакансий, большая часть их в Москве и Санкт-Петербурге. В DE нуждаются разные отрасли. В основном это продуктовые компании, ретейлеры, консалтинг, банки, телеком, производство, добыча сырья. Размер компаний зависит от отрасли, в основном это средние (около 100 человек) и большие компании (более 1000 человек).

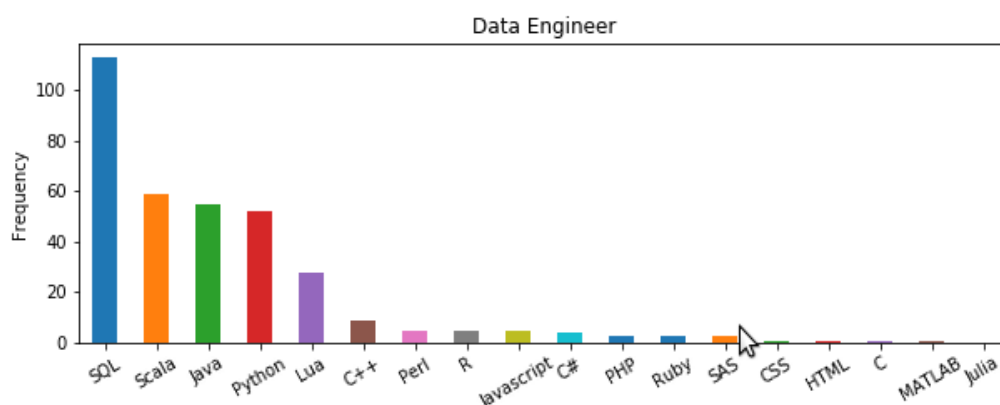
### 2) Требования: Популярные языки, технологии, фреймворки

Data engineer играет в организации критически важную роль – создает и поддерживает в рабочем состоянии пайплайны и базы данных, которые используются для обработки, трансформации и хранения данных. Для этого ему необходимы профессиональные следующие знания и навыки:

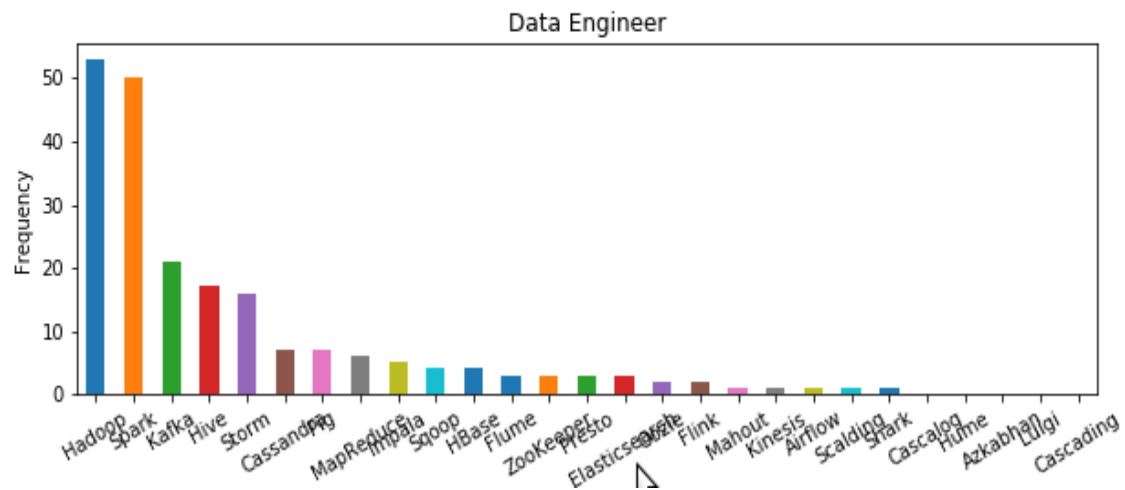
- алгоритмы и структуры данных;
- принципы хранения информации в SQL и NoSQL, а также умение работать с реляционными и нереляционными базами данных (MySQL, MSSQL, PostgreSQL, MongoDB, SQL Server, Oracle, HP Vertica, Amazon Redshift и т.д.)
- ETL-системы (Informatica, Pentaho ETL, Talend, Apache Airflow);
- облачные платформы для Big Data решений (Amazon Web Services, Google Cloud Platform, Microsoft Azure и другие подобные решения от крупных PaaS/IaaS-провайдеров);
- стек Apache Hadoop (HDFS, HBase, Cassandra) и SQL-движки для анализа данных, хранящихся в распределенных файловых системах типа HDFS (Apache Hive, Impala и пр.); кластеры Big Data на базе Apache (Hadoop, Kafka, Spark);
- языки программирования (Python, Java, Scala) для работы с Big Data системами.

### Технологии востребованные в Европе:

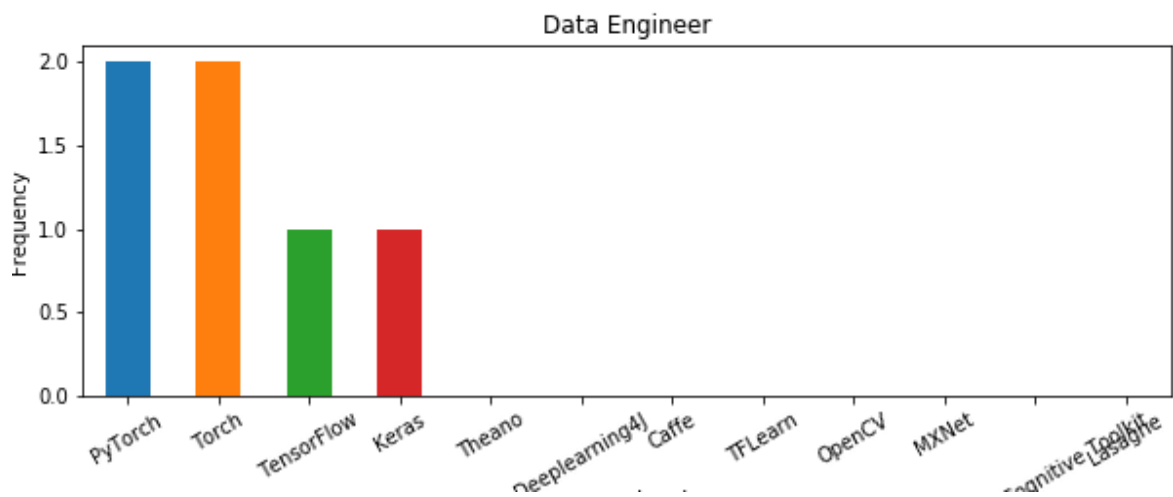
## Programming Languages Distribution



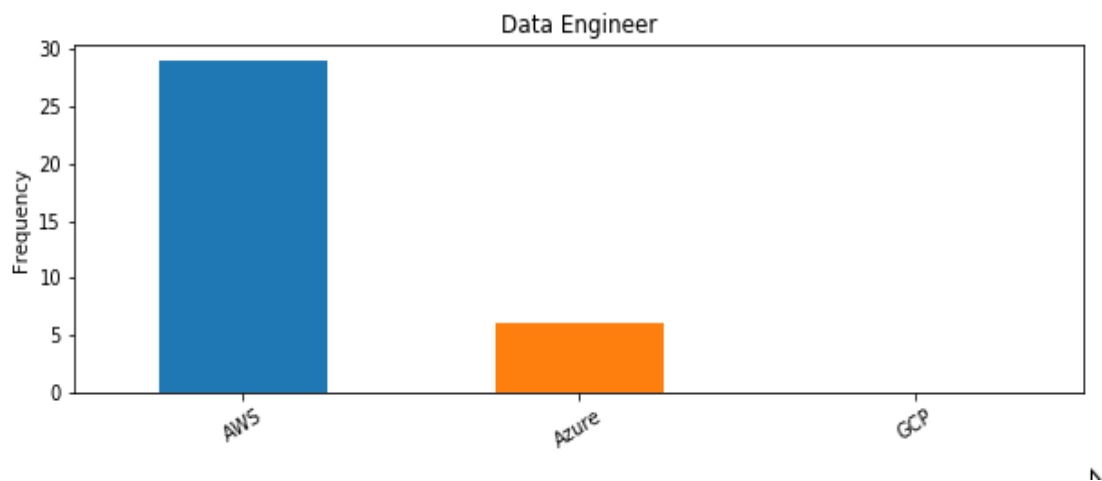
# Big Data Technologies Distribution



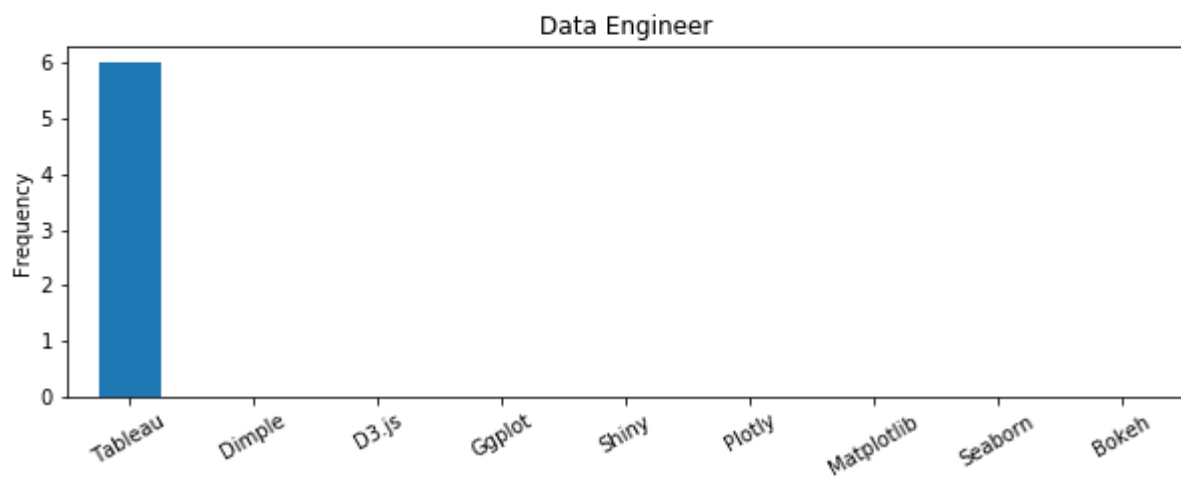
# Deep Learning Frameworks Distribution



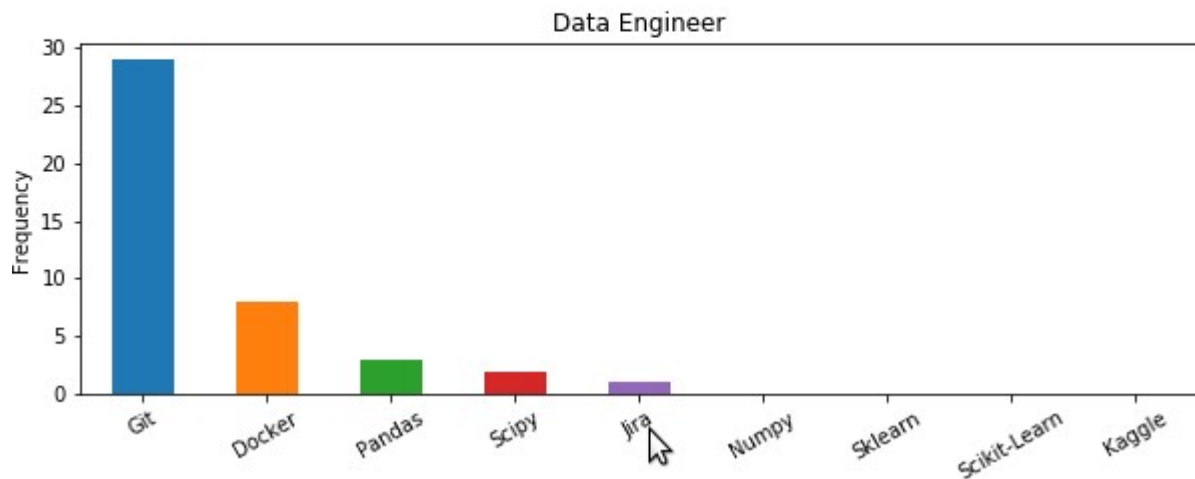
# Cloud Computing Platforms Distribution



# Visualization Tools Distribution



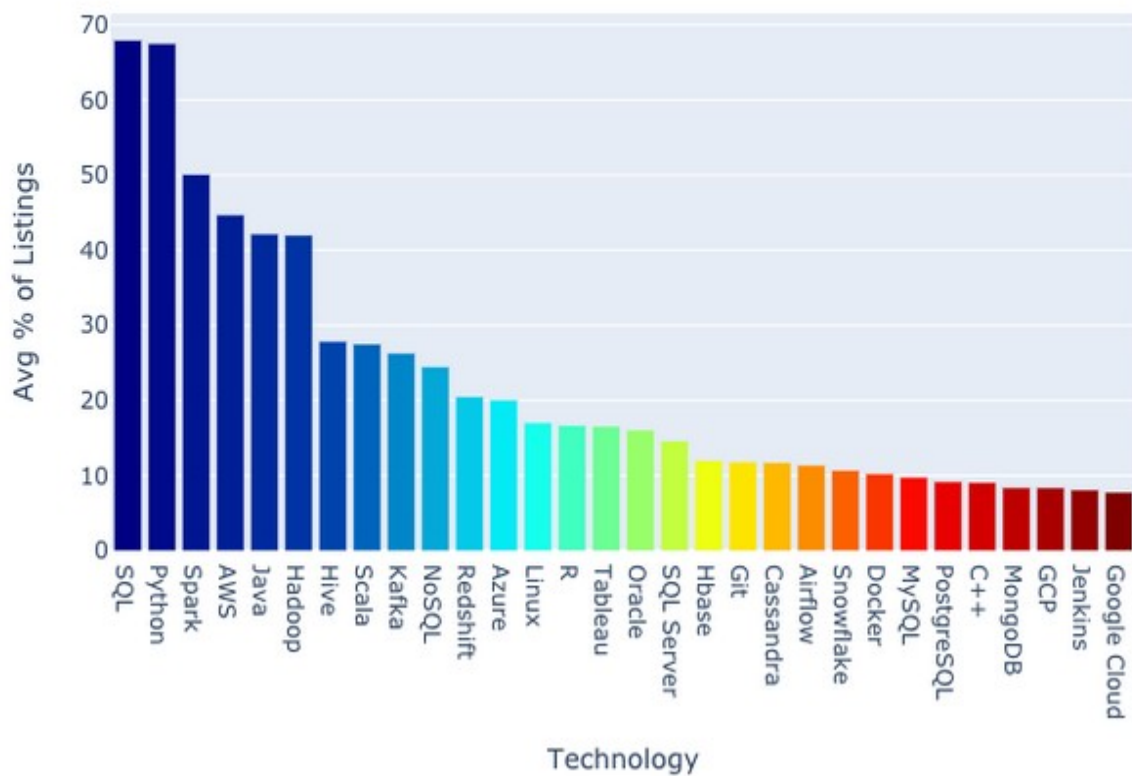
# Other Distribution



<https://elu.nl/what-does-an-ideal-data-scientists-profile-look-like/>

## Технологии востребованные в США:

### Technologies in Data Engineer Job Listings 2020



<https://towardsdatascience.com/most-in-demand-tech-skills-for-data-engineers-58f4c1ca25ab>

### **Технологии востребованные в РФ:**

Отличие в востребованных технологиях в РФ по сравнению с Европой и США в том, что в вакансиях в РФ знания Cloud не часто встречается.

### **3) Цели на обучение, акценты на интересующие темы и технологии**

Моя главная цель курса познакомиться со стеком Big Data. Так же мне интересны темы data governance и data quality. Получить практические навыки. Сделать выпускную работу используя стек Big Data . С темой я точно еще не определился, но хотел бы попробовать сделать рекомендательную систему используя Apache Spark.