**Анализ рынка труда по профессии инженера данных в России.**

1. По запросу Data Engineer / Data инженер hh.ru выдает 1136 вакансий по России,

по запросу инженер данных - 10590.

Отрасли компаний: финтех, телеком,ритейл. В основном крупные компании, работающие с большими объемами данных.

1. Стек технологий:

Языки: самые популярные языки Python и Scala, ~80% вакансий, также Java, Go, R.

БД: самые популярные БД - HBase, Cassandra, Postgres, OracleDB.

Самые востребованные технологии: Spark/Hadoop, Airflow (встречаются практически во всех вакансиях), Kafka, Hive, Flink, Datalake, Jenkins, ETL/ELT.

Опыт: вакансии без опыта - 5-6%, 1-3 года - 35-40%, 3-6 лет - 45-50%, более 6 лет - 8-10%.

1. Spark - фреймворк для реализации распределенной обработки больших данных. Самый популярный на данный момент. Можно работать на разных ЯП - Java, Python, Scala.

Hadoop - экосистема для разработки и выполнения распределенных программ, работающих на кластерах из сотен и тысяч узлов.

Airflow - фреймворк для создания, выполнения, мониторинга и управления потоков операций по обработке данных.

Kafka - распреленная горизонтально масштабируемая система. По сути брокер сообщений, работающий по принципу очереди. Часто используется для асинхронного взаимодействия между сервисами в приложении.

Numpy - библиотека Python, поддерживающая большие многомерные массивы и матрицы, высокоуровневые математические функции для операции с массивами.

Jenkins - программа/софт предназначенная для создания и поддержки процессов непрерывной интеграции и непрерывной доставки кода (CI/CD).

1. Получить компетенции в смежной для себя отрасли (java-разработчик), изучить новые технологии, перейти на новый проект в новой для себя роли, либо в качестве разработчика, но на проект связанный с большими данными. Меня интересуют фреймворки для реализации распределенной обработки данных(Spark, Hadoop), NoSQL базы данных, познакомиться с ML и нейронными сетями.