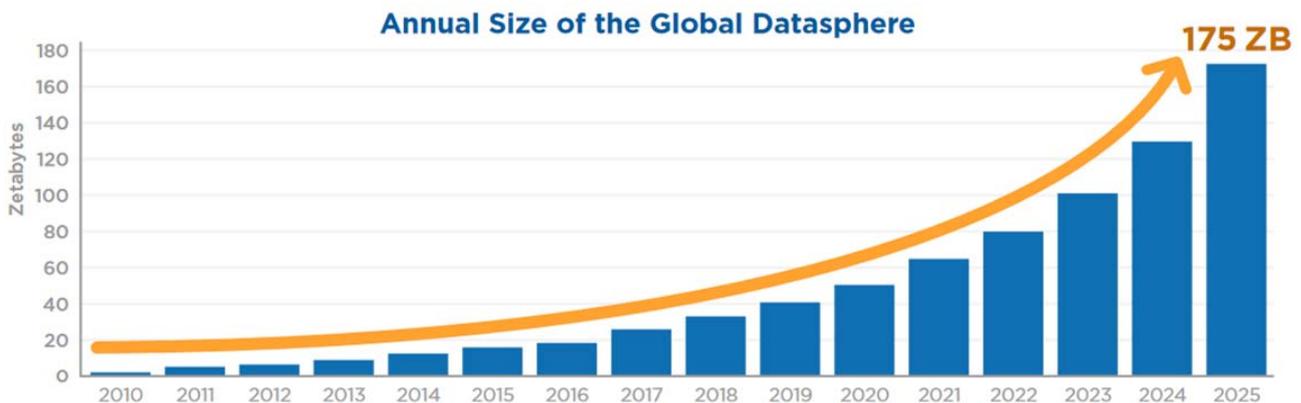


## ВИЗУАЛИЗАЦИЯ ДАННЫХ В PYTHON

Объем данных в мире постоянно и стремительно растет. В исследовании от International Data Corporation дается прогноз, что мировой объем данных с 33 зеттабайтов в 2018 г. вырастет до 175 зеттабайтов к 2025 г., где зеттабайт — это триллион гигабайт [1, с. 3]. Этот прогноз представлен на рисунке.



Ежегодный объем мировых данных

И с т о ч н и к: разработано автором на основе [1, с. 6].

С таким стремительным ростом объема данных необходимо еще более стремительное как интенсивное, так и экстенсивное развитие их анализа.

Для осознания актуальности визуализации данных и получения представления о том, как она используется в анализе данных, необходимо знать, что такое данные и что такое информация. Данные — это набор фактов в необработанном виде. Данные, из которых можно извлечь пользу, называются информацией. Чтобы получить информацию из данных, их нужно обработать и отсортировать. Визуализация данных — это процесс графического представления данных. Графическое представление данных ускоряет извлечение информации из них и, соответственно, ускоряет их анализ.

Целью работы является анализ возможностей для визуализации данных с помощью библиотек языка программирования Python, а также их сравнение с возможностями одного из самых популярных приложений для визуализации данных — Tableau.

Python — высокоуровневый язык программирования, получивший широкое распространение благодаря множеству многофункциональных библиотек. Наиболее востребованными библиотеками для визуализации данных являются

Matplotlib и Seaborn, совместно с которыми для анализа данных используются такие библиотеки, как NumPy и pandas. Tableau Desktop — приложение с большим функционалом и удобным графическим интерфейсом, специализирующееся на визуализации данных.

Основываясь на статье о возможностях визуализации данных в Python и Tableau Desktop [2], составим краткую таблицу их сравнения.

Сравнение Python и Tableau для визуализации данных

Критерий сравнения	Python	Tableau Desktop
Стоимость	Бесплатный	70 дол. в месяц
Удобство работы	Нужен опыт	Легок в освоении
Скорость работы	Зависит от опыта пользователя	Высокая
Возможности визуализации	Все типы графиков, возможность создания карт визуализаций	Все типы графиков, возможность создания карт визуализаций

Вывод: Tableau хорош простотой и удобством, но требует финансовых вложений и сфокусирован на визуализации данных, а Python сложнее в освоении, но имеет больше возможностей и бесплатен, а также имеет множество применений в различных сферах, что делает его более перспективным для организации полного цикла анализа данных — от сбора, очистки, загрузки до визуализации и аналитики.

### Источники

1. The Digitization of the World [Electronic resource] // Seagate. — Mode of access: <https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf>. — Date of access: 05.04.2020.
2. Can Python DataViz libraries reproduce the Tableau worksheet? [Электронный ресурс] // Towards Data Science. — Режим доступа: <https://towardsdatascience.com/can-python-dataviz-libraries-repeat-the-tableau-worksheet-e38ef2876f04>. — Дата доступа: 05.04.2020.