

# Построение графиков

Семинар 10

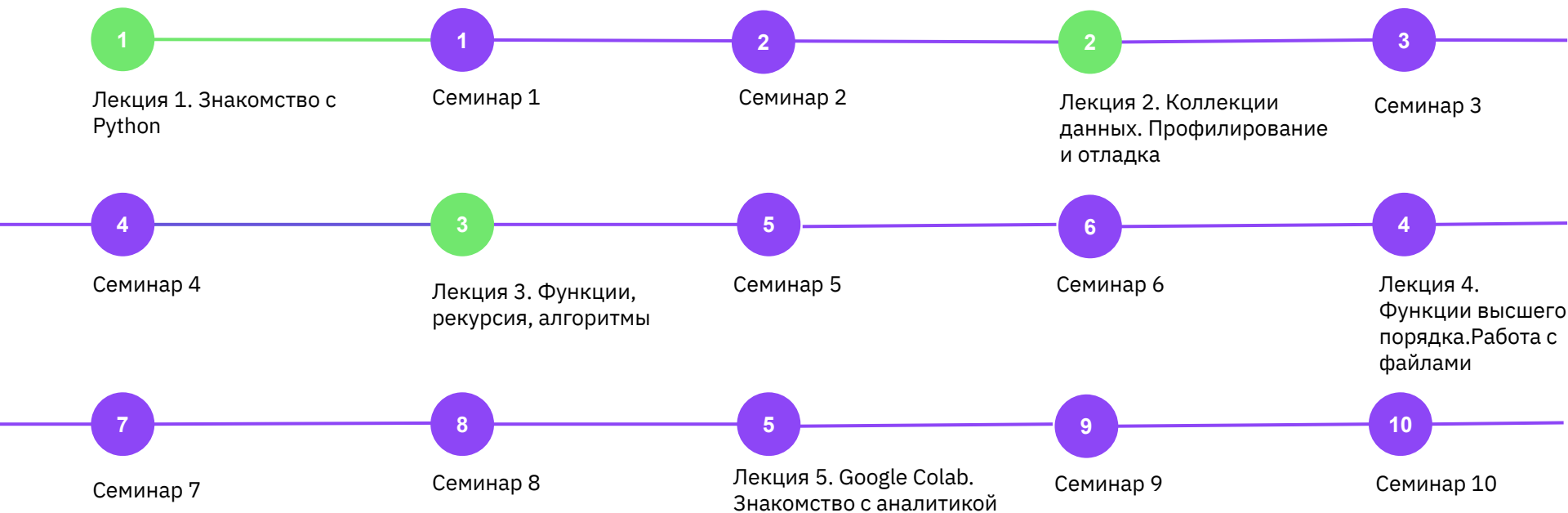




# Содержание урока








## План курса





## Что будет на уроке сегодня

-  Закрепить полученные знания с лекции по теме построение графиков
-  Обсуждение домашнего задания
-  Решение в группе задач
-  Домашнее задание
-  Подведение итогов

## Домашнее задание

### Задание

**Задача 40:** Работать с файлом california\_housing\_train.csv, который находится в папке sample\_data. Определить среднюю стоимость дома, где кол-во людей от 0 до 500 (population)

**Задача 42:** Узнать какая максимальная households в зоне минимального значения population





# Практика



## Задача №63. Решение в группах

1. Изобразите отношение households к population с помощью точечного графика
2. Визуализировать longitude по отношению к median house value, используя линейный график
3. Представить гистограмму по housing median age
4. Изобразить гистограмму по median house value с оттенком housing median age



20 минут



## Задача №63. Общее обсуждение

1. Изобразите отношение households к population с помощью точечного графика
2. Визуализировать longitude по отношению к median house value, используя линейный график
3. Представить гистограмму по housing median age
4. Изобразить гистограмму по median house value с оттенком housing median age



5 минут





## Задача №65. Решение в группах

Написать EDA для датасета про пингвинов

Необходимо:

- Использовать 2-3 точечных графика
- Применить доп измерение в точечных графиках, используя аргументы hue, size, style
- Использовать PairGrid с типом графика на ваш выбор
- Изобразить Heatmap
- Использовать 2-3 гистограммы

Чтобы подключить датасет с  
пингвинами, воспользуйтесь данным  
скриптом:

```
penguins = sns.load_dataset("penguins")  
penguins.head()
```



20 минут



## Задача №65. Общее обсуждение

Написать EDA для датасета про пингвинов

Необходимо:

- Использовать 2-3 точечных графика
- Применить доп измерение в точечных графиках, используя аргументы hue, size, style
- Использовать PairGrid с типом графика на ваш выбор
- Использовать 2-3 гистограммы

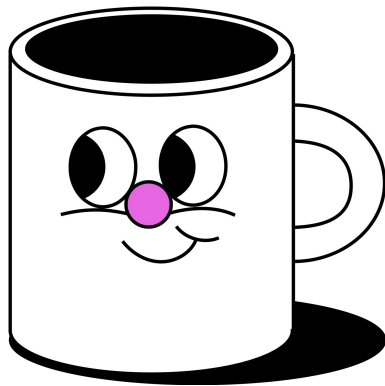
Чтобы подключить датасет с  
пингвинами, воспользуйтесь данным  
скриптом:

```
penguins = sns.load_dataset("penguins")  
penguins.head()
```



5 минут

**Перерыв**



**<<5:00->>**



## Задача №67. Решение в группах

1. Создать новый столбец в таблице с пингвинами, который будет отвечать за показатель длины клюва пингвина.  
high - длинный(от 42)  
middle - средний(от 35 до 42)  
low - маленький(до 35)



20 минут

Чтобы подключить датасет с пингвинами, воспользуйтесь данным скриптом:

```
penguins = sns.load_dataset("penguins")  
penguins.head()
```



## Задача №67. Общее обсуждение

1. Создать новый столбец в таблице с пингвинами, который будет отвечать за показатель длины клюва пингвина.  
high - длинный(от 42)  
middle - средний(от 35 до 42)  
low - маленький(до 35)



5 минут

Чтобы подключить датасет с пингвинами, воспользуйтесь данным скриптом:

```
penguins = sns.load_dataset("penguins")  
penguins.head()
```



## Задача №69. Решение в группах

1. Изобразить гистограмму по `flipper_length_mm` с оттенком `height_group`. Сделать анализ



20 минут



## Задача №69. Общее обсуждение

1. Изобразить гистограмму по flipper\_length\_mm с оттенком height\_group. Сделать анализ



**5 минут**



Вопросы?

Вопросы?



Вопросы?







# Домашнее задание

## Задание

**Задача 44:** В ячейке ниже представлен код генерирующий *DataFrame*, которая состоит всего из 1 столбца. Ваша задача перевести его в *one hot* вид. Сможете ли вы это сделать без *get dummies*?

```
import random
lst = ['robot'] * 10
lst += ['human'] * 10
random.shuffle(lst)
data = pd.DataFrame({'whoAmI':lst})
data.head()
```

[Статья про one hot вид](#)





## Рефлексия



**Был урок полезен вам?**



**Узнали вы что-то новое?**



**Что было сложно?**



Спасибо за внимание!