Цель проекта:

1:научиться делать математическую модель и проверить гепотизу

2:повторить библеотеку pandas

1 ШАГ-

проверить свою придуманную гипотезу

```python
df = pd.read_csv('train.csv')
df['langs'].fillna(-1, inplace = True)
df['result'].fillna(-1, inplace = True)
langs_count = 0
langs_rating = 0
def langs_avarage(row):
    global langs_count , langs_rating
    if len(row['langs'].split(';'))> 1:
        langs_count += row['result']
        langs_rating += 1
    else:
        langs_count += row['result']
        langs_rating += 0.1
    return False

df.apply(langs_avarage, axis = 1)
ss = pd.Series(data = [langs_count , langs_rating], index = ['языковеды', 'обычные'
ss.plot(kind = 'pie', label = ' ')
plt.show()
```

2 шаг-

создание математической модели

```
x =df.drop(['result'], axis =1)
y = df['result']
x_train , x_test , y_train , y_test  =train_test_split(x, y , test_size = 0.90)
sc = StandardScaler()
x_train = sc.fit_transform(x_train)
x_test =sc.transform(x_test)
classifier = KNeighborsClassifier(n_neighbors = 5)
classifier.fit(x_train, y_train)

y_pred = classifier.predict(x_test)
print(y_test)
print(y_pred)
print('процент правильного рпедсказания исхода :', round(accuracy_score(y_test, y_pred)*100, 2))
```

3 шаг-

ОТКАРИКТИРОВАТЬ ВСЕ СТОЛБЦЫ

```
def sex_apply(sex):
    if sex ==2:
        return 0
    return 1

df['education_form'].fillna('Full-time', inplace = True)
df[list(pd.get_dummies(df['education_form']).columns)] = pd.get_dummies(df['education_form'])
df.drop(['education_form'], axis =1,  inplace =True)


def edu_status_apply(education_status):
    if education_status == "Undergraduate applicant":
        return 0
    elif education_status == "Student (Master's)"  or education_status == "Student (Bachelor's) " or education_status ==
        return 1
    elif education_status == "Alumnus (Master's)" or education_status  == "Alumnus (Bachelor's)" or education_status ==
        return 2
    else:
        return 3
df['education_status'] = df['education_status'].apply(edu_status_apply)

def langs_apply(langs):
    if langs.find('Русский') != 1 and langs.find('English'):
        return 0
    return 1
df['langs'] = df['langs'].apply(langs_apply)

def occupation_type_apply(occupation_type):
    if occupation_type == "university":
        return 0
    return 1
df['occupation_type'] = df['occupation_type'].apply(occupation_type_apply)
df.info()
```

ВЫВОД:

-выполнено

тема урока было очень интересная и пользовательная

СПАСИБО ЗА ВНИМПНИЕ

OFFICE

- проверить гипотезу
- создать математическую модель
- откариктировать столбцы