**ШИНЖЛЭХ УХААН ТЕХНОЛОГИЙН ИХ СУРГУУЛЬ**

**Мэдээлэл холбооны технологийн сургууль**

****

# **Их өгөгдлийн шинжилгээ**

**30-н оноо**

**2022-2023 оны хичээлийн жилийн хавар**

**Лабораторийн цаг :** 5-2

**Хянасан багш:** Г.Мөнхбат

**Гүйцэтгэсэн:**   **Б**. Эрдэнэ-Мандал /B201900001/

**Э**. Тэнгис /B201900002/

**Улаанбаатар хот**

**2023 он**

**Даалгавар:** Attention visualization Neftlix

**Хураангуй**

Энэхүү машин сургалтын гол зорилго бол neftlix дээрх киноны үзэгчид болон ямар кино хит болж байгааг ажиглах юм.

**Зорилго**

Анхаарал татахуйц дүрслэл нь машин сургалтын загвар хэрхэн ажилладагийг ойлгох хүчирхэг хэрэгсэл юм. Загварын таамаглалд хамгийн чухал байгаа оролтын хэсгүүдийг тодорхойлоход үүнийг ашиглаж болно. Энэ нь загваруудыг дибаг хийх, загваруудын гүйцэтгэлийг сайжруулахад тустай байж болно.

**Attention visualization гэж юу вэ? Архитектур ба горимууд**

Анхаарал татахуйц дүрслэл нь машин сургалтын загвар нь оролтын өөр өөр хэсгүүдэд хэрхэн оролцож байгааг ойлгоход ашигладаг арга юм. Үүнийг машин орчуулга, текстийн хураангуй, зургийн тайлбар зэрэг олон төрлийн ажлуудад хийж болно.

Анхаарал хандуулах олон янзын арга байдаг. Нэг нийтлэг арга бол анхаарлын жинг дулааны зураг болгон зурах явдал юм. Энэ нь оролтын хэсэг бүр хэр их анхаарал хандуулж байгааг харуулдаг бөгөөд бараан өнгө нь илүү анхаарал хандуулж байгааг харуулж байна.

Өөр нэг арга бол цаг хугацааны дүрслэлийг ашиглах явдал юм. Энэ нь загвар нь оролтыг боловсруулах явцад анхаарлын жин цаг хугацааны явцад хэрхэн өөрчлөгдөж байгааг харуулж байна.

Анхаарал татахуйц дүрслэл нь машин сургалтын загвар хэрхэн ажилладагийг ойлгоход хэрэгтэй хэрэгсэл байж болно. Энэ нь загварын таамаглалд хамгийн чухал оролтын хэсгүүдийг тодорхойлоход тусална. Энэ нь загваруудыг дибаг хийх, загваруудын гүйцэтгэлийг сайжруулахад тустай байж болно.

Анхаарал татахуйц дүрслэлийн зарим жишээ энд байна:

Машины орчуулгад загвар нь эх хэл дээрх үгсийг зорилтот хэл дээрх үгстэй хэрхэн нийцүүлж байгааг харахын тулд анхаарлын дүрслэлийг ашиглаж болно. Энэ нь загвар нь текстийг хэрхэн орчуулж байгааг ойлгоход тусална.

Текстийг нэгтгэн дүгнэхэд загвар нь оролтын текстээс хамгийн чухал өгүүлбэрүүдийг хэрхэн сонгож байгааг харахын тулд анхаарлын дүрслэлийг ашиглаж болно. Энэ нь загвар хураангуйг хэрхэн үүсгэж байгааг ойлгоход тусална.

Зургийн тайлбарт загвар нь зургийн өөр өөр хэсгүүдэд хэрхэн анхаарал хандуулж байгааг харахын тулд анхаарлын дүрслэлийг ашиглаж болно. Энэ нь загвар нь тайлбарыг хэрхэн үүсгэж байгааг ойлгоход тусална.

Анхаарал хандуулахад ашиглаж болох зарим хэрэгслүүд энд байна:

BertViz: BertViz нь BERT, GPT2, T5 гэх мэт Transformer хэлний загварт анхаарлаа хандуулах интерактив хэрэгсэл юм.

Tensor2Tensor: Tensor2Tensor нь дулааны зураглал, цагийн хуваарийн дүрслэл зэрэг анхаарлыг дүрслэн харуулах олон хэрэгслээр хангадаг номын сан юм.

nnExplainer: nnExplainer нь машин сургалтын загваруудын таамаглалыг тайлбарлах олон хэрэгслээр хангадаг номын сан бөгөөд үүнд анхаарлын дүрслэлийг оруулав.

Анхаарал татахуйц дүрслэл нь машин сургалтын загвар хэрхэн ажилладагийг ойлгох хүчирхэг хэрэгсэл юм. Загварын таамаглалд хамгийн чухал байгаа оролтын хэсгүүдийг тодорхойлоход үүнийг ашиглаж болно. Энэ нь загваруудыг дибаг хийх, загваруудын гүйцэтгэлийг сайжруулахад тустай байж болно.

**What is seaborn**

Seaborn бол matplotlib дээр суурилсан Python дүрслэх номын сан юм. Энэ нь сонирхол татахуйц, мэдээлэл сайтай статистик график зурах өндөр түвшний интерфейсээр хангадаг.

Seaborn нь matplotlib дээр бүтээгдсэн боловч үзэсгэлэнтэй, мэдээлэл сайтай дүрслэлийг бий болгоход хялбар болгодог хэд хэдэн функцээр хангадаг. Жишээлбэл, Seaborn нь өөрийн талбайн үзэмж, мэдрэмжийг өөрчлөхөд ашиглаж болох хэд хэдэн урьдчилан тодорхойлсон хэв маягийг санал болгодог. Seaborn нь мөн тархсан график, шугамын график, дулааны зураг зэрэг нийтлэг статистик график үүсгэхэд хялбар болгодог хэд хэдэн функцээр хангадаг.

Seaborn бол өгөгдлийг дүрслэх хүчирхэг хэрэгсэл бөгөөд үүнийг олон төрлийн зураглал үүсгэхэд ашиглаж болно. Seaborn нь янз бүрийн хувьсагчдын хоорондын хамаарлыг харуулсан зураглал үүсгэхэд онцгой ач холбогдолтой юм. Жишээлбэл, Seaborn-ийг ашиглан байшингийн үнэ болон түүний квадрат метр хоорондын хамаарал, эсвэл ангийн сурагчдын тоо болон шалгалтын онооны хоорондын хамаарлыг харуулсан зураглал үүсгэж болно.

Seaborn бол өгөгдлийг дүрслэн харуулах алдартай номын сан бөгөөд үүнийг дэлхийн өнцөг булан бүрт өгөгдөл судлаачид, шинжээчид ашигладаг. Хэрэв та үзэсгэлэнтэй, мэдээлэл сайтай статистик график бүтээх арга хайж байгаа бол Seaborn бол маш сайн сонголт юм.

Seaborn-ийн зарим онцлог шинж чанарууд энд байна.

Урьдчилан тодорхойлсон хэв маяг: Seaborn нь өөрийн талбайн үзэмж, мэдрэмжийг өөрчлөхөд ашиглаж болох хэд хэдэн урьдчилан тодорхойлсон хэв маягийг санал болгодог.

Нийтлэг статистикийн графикийн функцууд: Seaborn нь тараах график, шугамын график, дулааны зураг зэрэг нийтлэг статистикийн график үүсгэхэд хялбар болгодог хэд хэдэн функцээр хангадаг.

Pandas-тай нэгтгэх: Seaborn нь Pandas-тай нэгддэг бөгөөд энэ нь Pandas DataFrames-ээс зураг үүсгэхэд хялбар болгодог.

Хэрэглэхэд хялбар: Seaborn нь эхлэгчдэд ч хэрэглэхэд хялбар байдаг.

Seaborn-ийн ашиг тусыг энд дурдав.

Seaborn нь танд үзэсгэлэнтэй, мэдээлэл сайтай статистик график бүтээхэд тусална.

Seaborn танд олж мэдсэн зүйлээ бусдад илүү үр дүнтэй хүргэхэд тусална.

Seaborn нь таны өгөгдөл дэх хэв маягийг тодорхойлоход тусална.

Seaborn нь таны өгөгдөл дээр үндэслэн илүү сайн шийдвэр гаргахад тусална.

Хэрэв та үзэсгэлэнтэй, мэдээлэл сайтай статистик график бүтээх арга хайж байгаа бол Seaborn бол маш сайн сонголт юм.

**What is scikit sklearn**

Sklearn.preprocessing import StandardScaler болон sklearn.cluster-аас импортлох KMeans, AffinityPropagation-аас импортын мэдэгдлүүд нь scikit-learn номын сангаас дараах классуудыг импортолдог:

StandardScaler: Өгөгдлийн багц дахь функцуудыг стандартчилахад ашиглаж болох анги. Энэ нь функцуудыг дундаж утга нь 0, стандарт хазайлт нь 1 байхаар томруулна гэсэн үг. Энэ нь онцлог шинж чанаруудын хэмжээнд мэдрэмтгий машин сургалтын алгоритмуудад тустай байж болно.

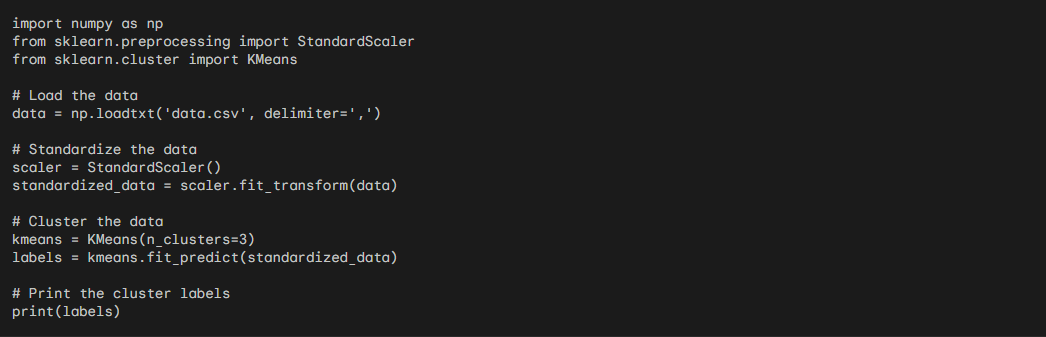
KMeans: Өгөгдлийн цэгүүдийг тодорхой тооны бүлэгт нэгтгэхэд ашиглаж болох анги. Алгоритм нь бүлгүүдийн центроидуудыг олж, дараа нь өгөгдлийн цэг бүрийг хамгийн ойр төвтэй бүлэгт оноох замаар ажилладаг.

AffinityPropagation: Өгөгдлийн цэгүүдийг ижил төстэй байдалд нь үндэслэн кластерлахад ашиглаж болох анги. Алгоритм нь хоорондоо хамгийн төстэй хос өгөгдлийн цэгүүдийг олж, дараа нь эдгээр хосуудын эргэн тойронд кластер үүсгэх замаар ажилладаг.

Эдгээр ангиудыг өгөгдлийг урьдчилан боловсруулах, кластер хийх олон төрлийн ажлыг гүйцэтгэхэд ашиглаж болно. Жишээлбэл, StandardScaler ангийг машин сургалтын алгоритмаар ашиглахын өмнө функцуудыг стандартчилахад ашиглаж болно. KMeans классыг өгөгдлийн цэгүүдийг тодорхой тооны бүлэгт нэгтгэхэд ашиглаж болно. AffinityPropagation классыг ижил төстэй байдалд үндэслэн өгөгдлийн цэгүүдийг кластерлахад ашиглаж болно.

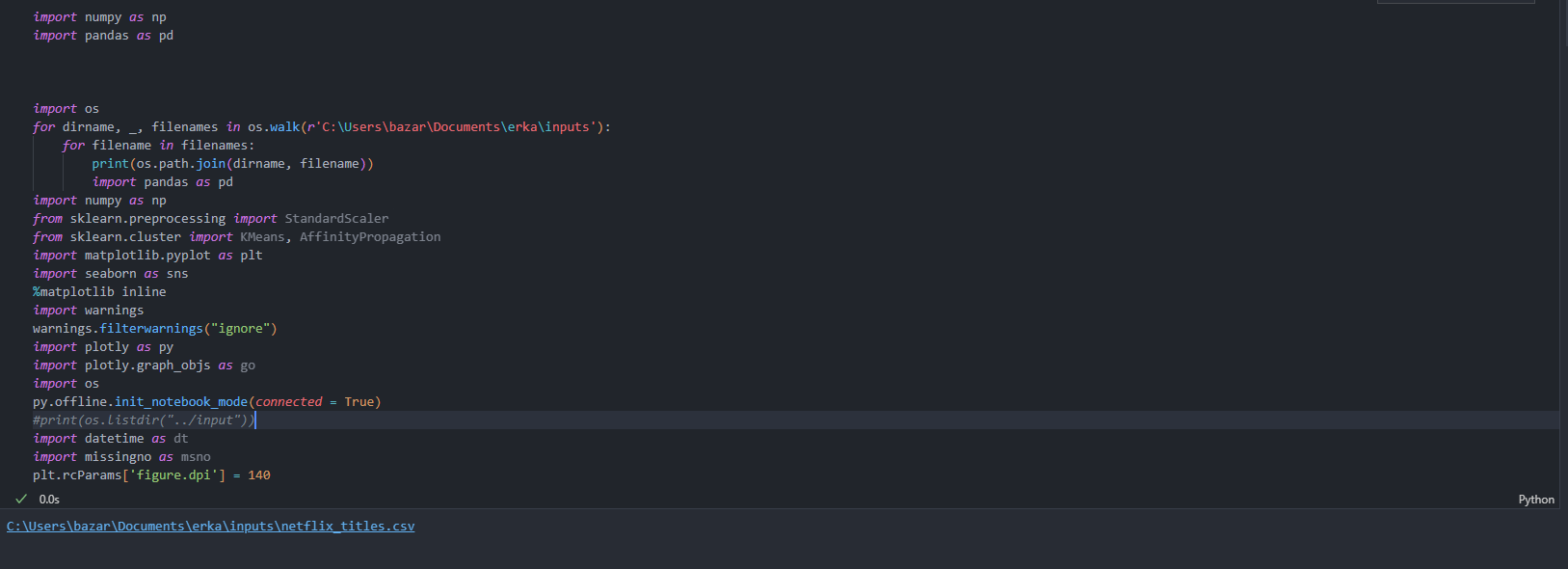
Эдгээр ангиудыг хэрхэн ашиглаж болох жишээ энд байна:

Кодын хэсэг



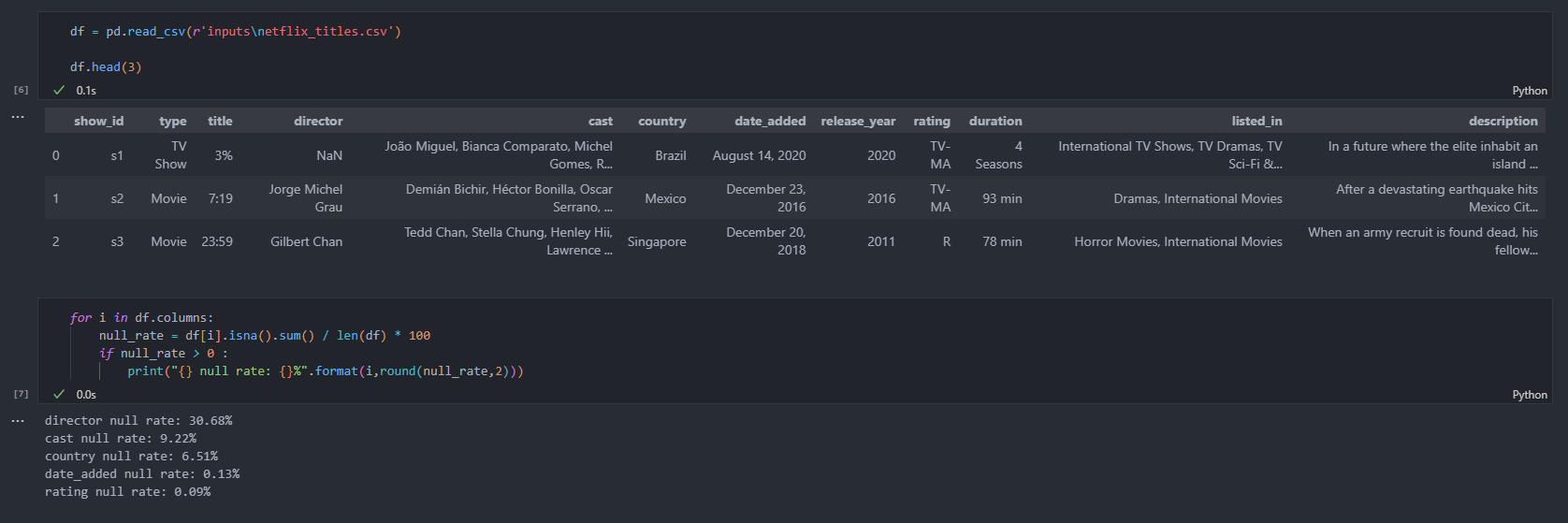
Энэ код нь data.csv файлаас өгөгдлийг ачаалж дараа нь өгөгдлийг стандартчилна. Дараа нь KMeans алгоритмыг ашиглан өгөгдлийг гурван бүлэгт нэгтгэнэ. Дараа нь кластерын шошго нь консол дээр хэвлэгдэх болно.

**Кодын хэсэг**



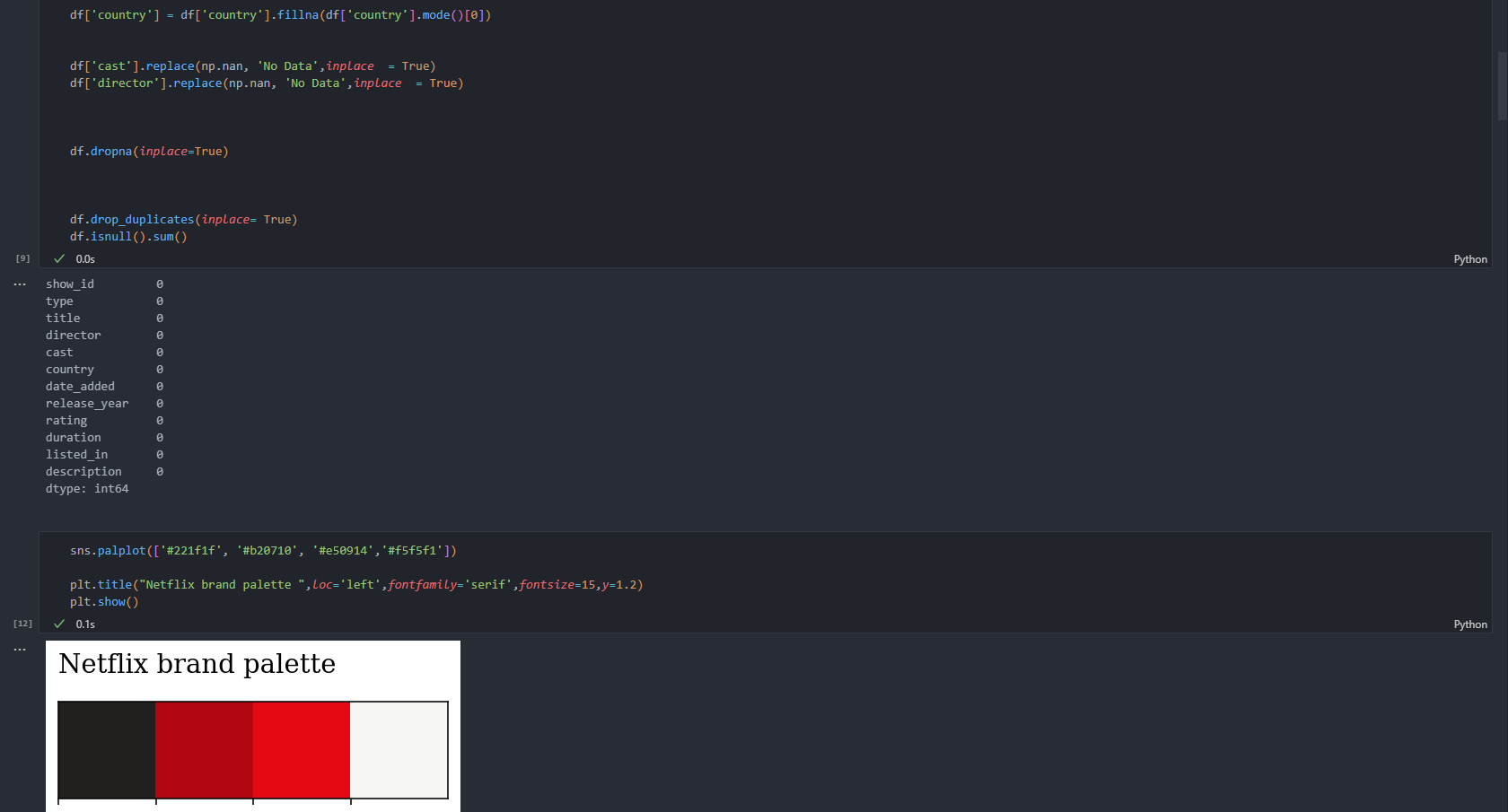
Хэрэгтэй сангуудаа импорт хийж оруулж ирнэ

**Алхам 2**

****

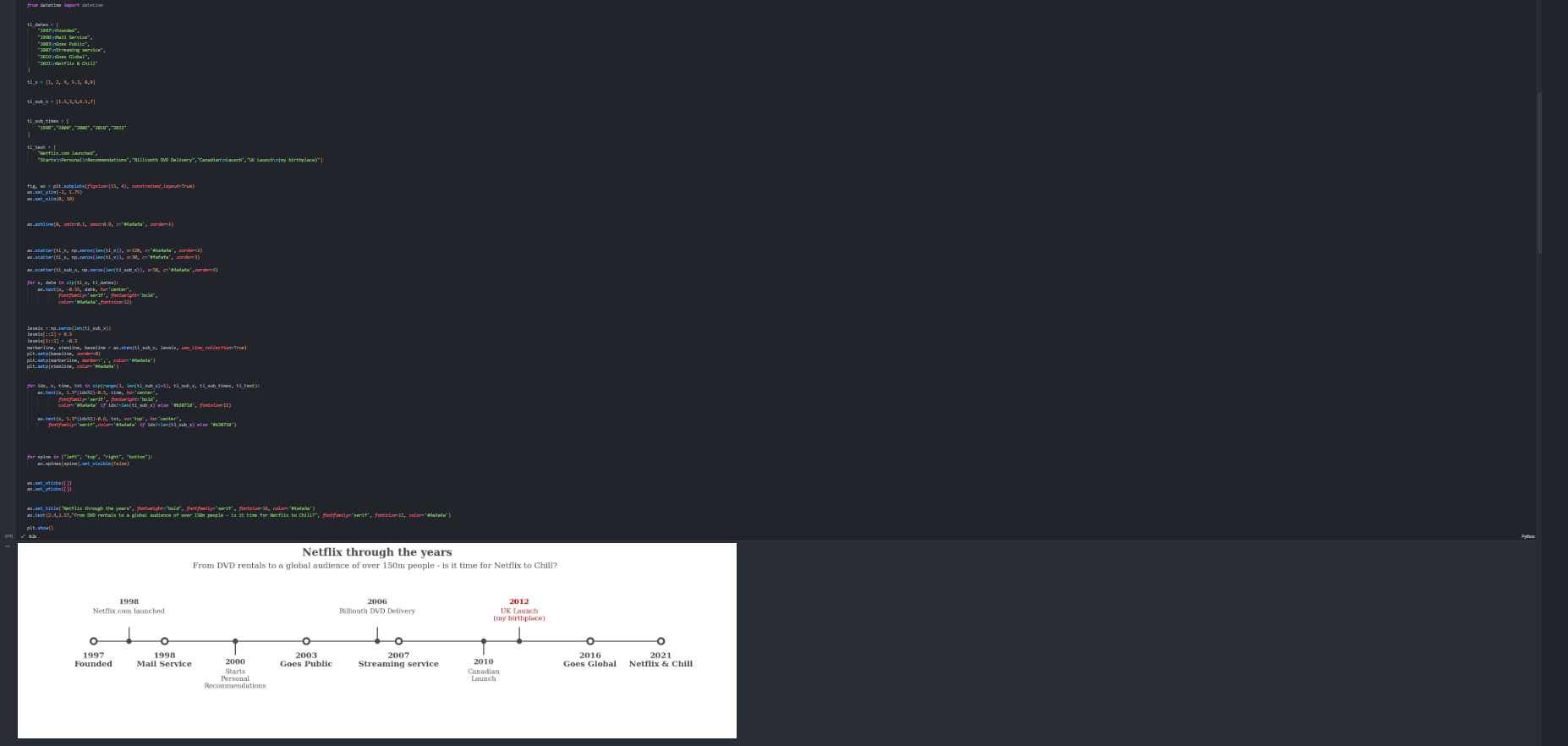
Өгөгдөлөө уншиж өгөгдлийн багана дундаа давталж ашиглаж нэрнүүдийн гаргаж авсан.

**Алхам 3**



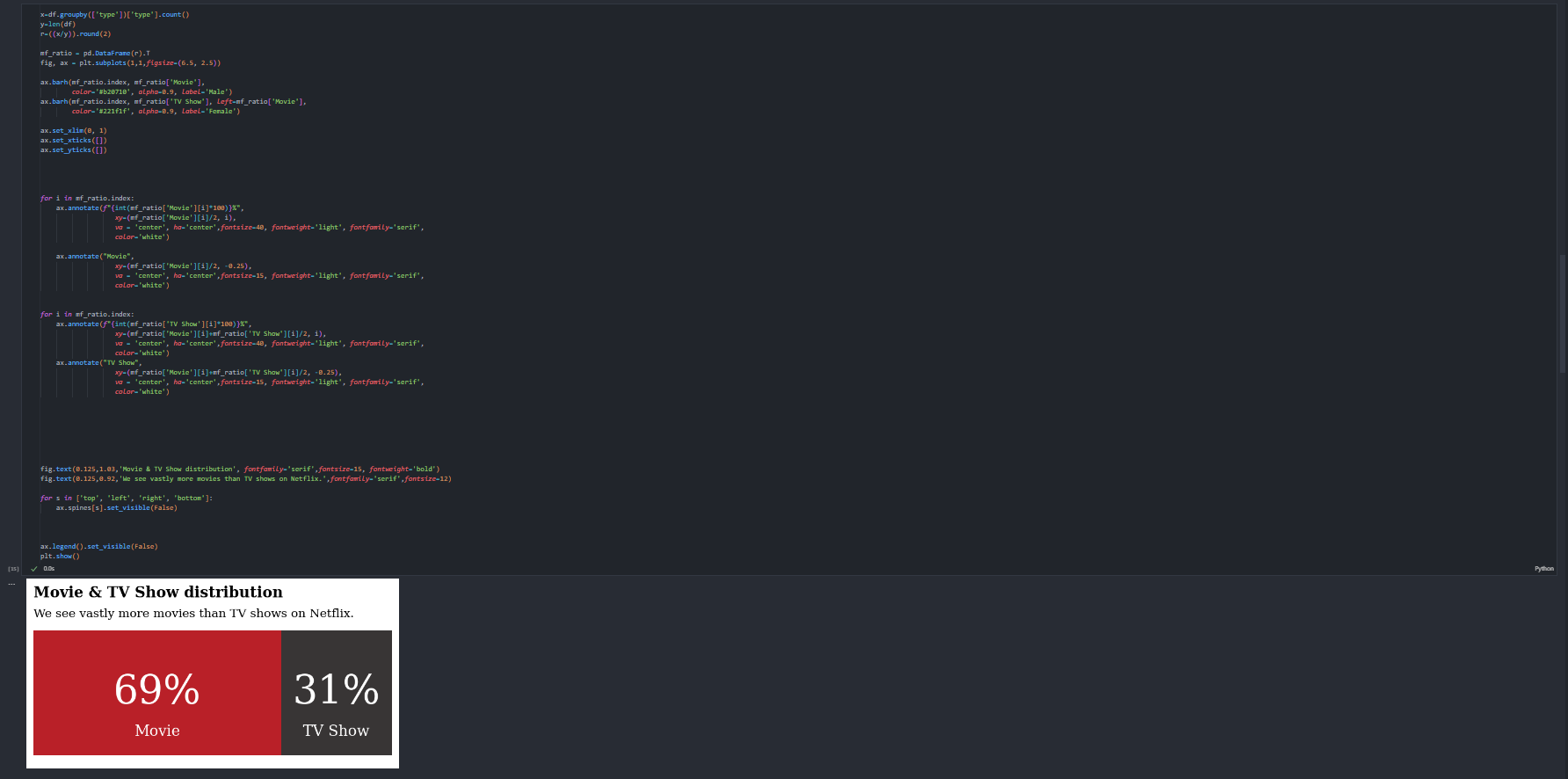
Дараа нь гаргаж авсан багана дунд хэрэгтэй баганаа түүвэрлэж seaborn сангийн sns ээр график гарганаавсан

**Алхам 4**

****

Энэ код дээр Datetime санг оруулж хамгийн их үзэлттэй хит болсон топ киног дүрсэлж харуулсан

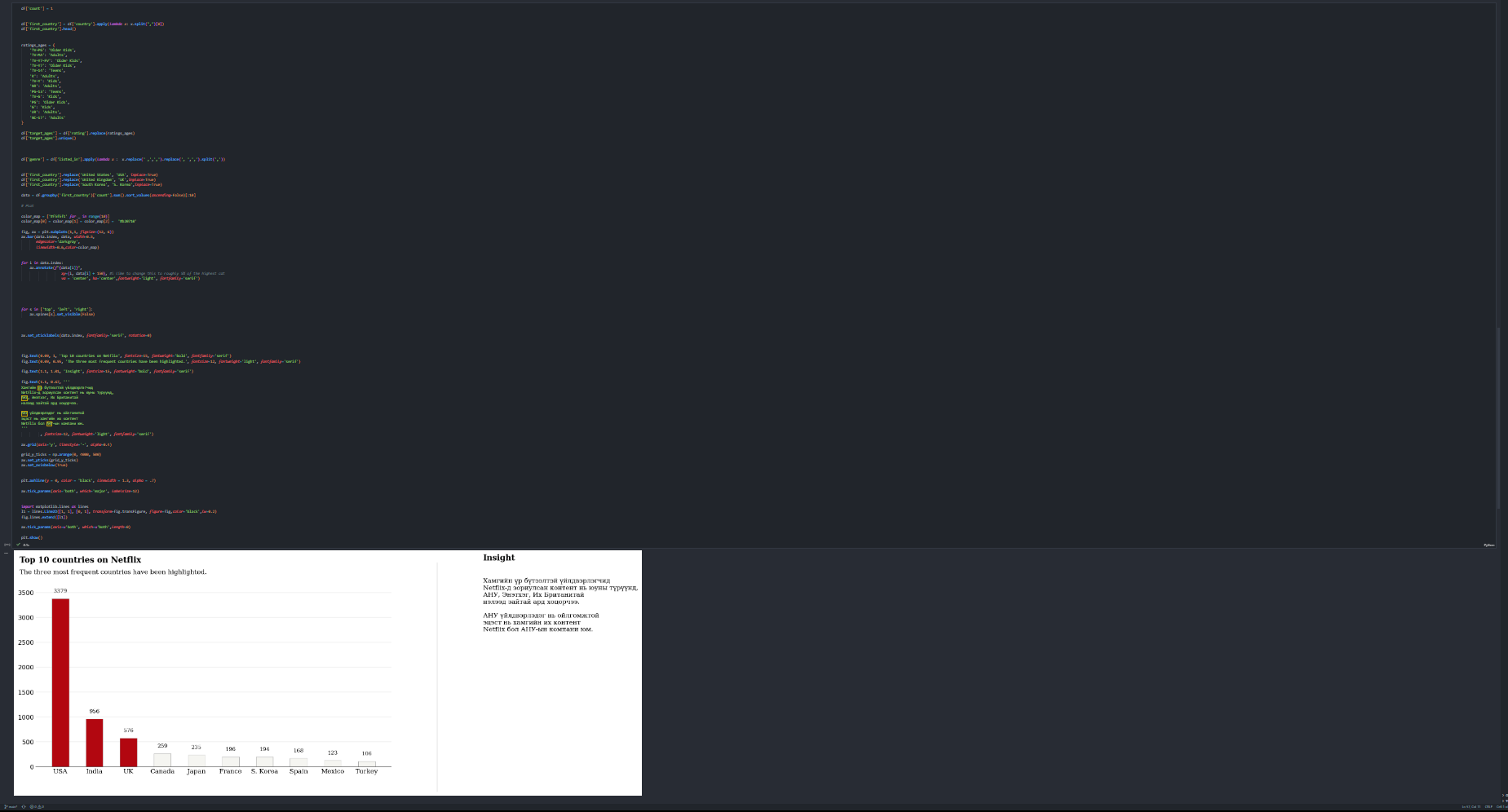
**Алхам 5**



Харин энэхүү хэсгээс үзэгчид ихэвчлэн Neftlix-ээс юу үздэг болохыг харж болно

69 хувь нь кино үздэг бол үлдсэн 31 хувь TV Show үздэг гэсэн судалгааг гаргаж авсан

**Алхам 6**

****

Хамгийн их үзэллтэй контентийг аль улс хийсэн гэдгийг дүрсэлж гаргаж ирсэн