

# Report - zima

## Osadský: Exploratory data analysis

V rámci ročníkového projektu sme v zimnom semestri spravili nasledovné:

### 1. Inštalácia virtuálneho prostredia Linux:

- Ako prvý krok sme úspešne spustili virtuálne prostredie Linux na počítači. Toto prostredie bolo kľúčové pre ďalší postup našej práce.

### 2. Spustenie anotačného nástroja pre CNV (Copy Number Variation):

- Na virtuálnej mašine sme úspešne spustili [anotačný nástroj pre CNV s chromozómami](#). Tento krok bol nevyhnutný pre získanie anotovaných dát pre ďalšiu analýzu. Tento krok predstavuje dôležitú etapu pre ďalšiu manipuláciu s dátami, keďže detailne opisuje vlastnosti chromozómu, ktoré skúmame. Anotované dáta v JSON formáte sme spracovali a prichystali sme ich do príslušného adresára.

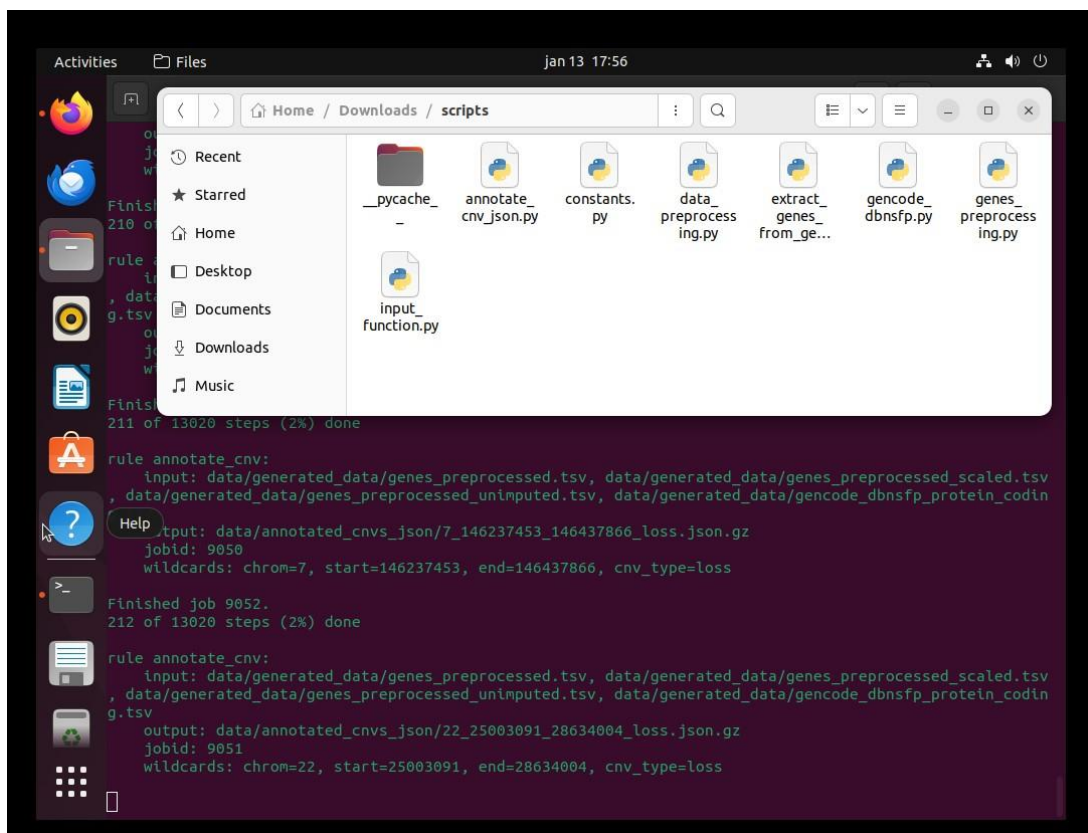


Figure 1 - annotation process

### 3. Spustenie Jupyter prostredia:

- Úspešne sme spustili Jupyter prostredie, kde sme následne pracovali s python notebookom. Toto prostredie nám poskytlo pohodlné a interaktívne možnosti práce s dátami dôležité pre ďalší postup, keďže vykresluje príkazy v pythone priebežne, takže nám tabuľku ukáže aj „kuse kódu.“

#### 4. Štúdium cvičení z exploratívnej analýzy dát:

- Aby sme lepšie porozumeli analýze dát, venovali sme sa štúdiu cvičení z oblasti [exploratívnej analýzy](#). Tento krok nám poskytol teoretický základ pre ďalšie spracovanie dát a poskytol úvod do knižnice Pandas.

#### 5. Vytvorenie základnej analýzy pomocou Pandas a Numpy:

- V prostredí Jupyter sme použili knižnice Pandas a Numpy na vytvorenie základnej analýzy dát. Práca s týmito knižnicami nám umožnila efektívne spracovanie a manipuláciu s dátami. Vygenerovali sme histogramy pre vybrané atribúty, čo predstavuje prvý krok pri vizuálnej analýze.

	chrom	start	end	cnv_type_x	genes_x	cnv_type_y	genes_y
0	10	100194214	132432797	gain	[{'Pathway(ConsensusPathDB)': 'imputed_unscal...	NaN	NaN
1	10	100600491	133622588	gain	[{'Pathway(ConsensusPathDB)': 'imputed_unscal...	NaN	NaN
2	10	100677716	100764519	loss	[{'Pathway(ConsensusPathDB)': 'imputed_unscal...	NaN	NaN
3	10	101120346	101831908	gain	[{'Pathway(ConsensusPathDB)': 'imputed_unscal...	NaN	NaN
4	10	101177304	101719109	gain	[{'Pathway(ConsensusPathDB)': 'imputed_unscal...	NaN	NaN
5	10	101209581	101748381	gain	[{'Pathway(ConsensusPathDB)': 'imputed_unscal...	NaN	NaN
6	10	101253668	101395093	gain	[{'Pathway(ConsensusPathDB)': 'imputed_unscal...	NaN	NaN
7	10	101261618	101468818	gain	[{'Pathway(ConsensusPathDB)': 'imputed_unscal...	NaN	NaN
8	10	101297762	101618190	gain	[{'Pathway(ConsensusPathDB)': 'imputed_unscal...	NaN	NaN
9	10	101370486	101583324	gain	[{'Pathway(ConsensusPathDB)': 'imputed_unscal...	gain	[{'Pathway(ConsensusPathDB)': 'imputed_unscal...

Figure 2 - table of annotated data

Celkovo vzaté, naša práca zahŕňala úspešnú inštaláciu, spustenie nástrojov na anotáciu a následné spracovanie anotovaných dát v prostredí Jupyter. Štúdium teoretických základov exploratívnej analýzy nám umožnilo lepšie porozumieť procesom analýzy dát, ktoré sme následne úspešne aplikovali pri vytváraní základnej analýzy s využitím Pandas a Numpy.