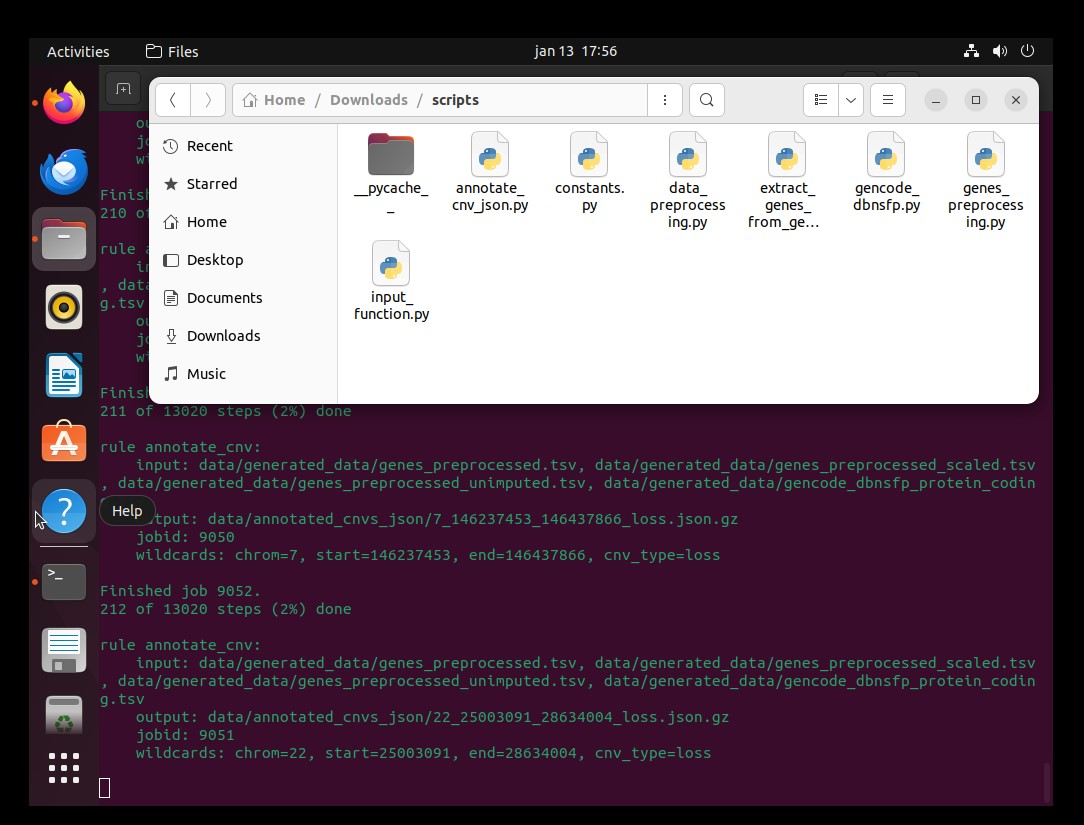
**Report - zima**

**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](https://github.com/tsladecek/cnv_gene_annotation/tree/master). 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.



1. **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.“
2. **Štúdium cvičení z exploratívnej analýzy dát:**
   * Aby sme lepšie porozumeli analýze dát, venovali sme sa štúdiu cvičení z oblasti [exploratívnej analýzy](https://github.com/FIIT-IAU/IAU-course/tree/main/exercises). Tento krok nám poskytol teoretický základ pre ďalšie spracovanie dát a poskytol úvod do knižnice Pandas.
3. **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.



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.