# Projekt Alapító Dokumentum (PAD)

Project: Project\_01

Dátum: [Helyezd be a dátumot]

## 1. Projekt adatai

Projekt neve: Project\_01

Projekt kezdete: [Helyezd be a kezdés dátumát]

Projekt vége: [Helyezd be a befejezés dátumát]

Projektvezető: [Helyezd be a projektvezető nevét]

Megrendelő: [Helyezd be a megrendelő nevét]

## 2. Projekt célja

A Project\_01 célja, hogy egy hatékony adatfeldolgozó és elemző rendszert hozzon létre, amely különféle adatelemzési és vizualizációs funkciókat valósít meg a Python programozási nyelv segítségével. A projekt célja, hogy felhasználóbarát, könnyen karbantartható és bővíthető legyen, amely különösen alkalmas oktatási és kutatási célokra.

## 3. Projekt terjedelme

A projekt magában foglalja:  
- Adatfeldolgozó funkciók implementálása, beleértve az adatok tisztítását és előfeldolgozását.  
- Statisztikai elemzések és gépi tanulási algoritmusok megvalósítása.  
- Grafikonok és vizualizációk készítése a matplotlib és más vizualizációs könyvtárak használatával.  
- Adatbetöltés különböző forrásokból, mint például CSV-fájlok.  
- Dokumentáció és oktatási anyagok biztosítása a felhasználóknak.

## 4. Felsőszintű követelmények

A projekt fő követelményei a következők:  
- Hatékony adatbetöltés és tisztítás.  
- Pontos statisztikai elemzések és vizualizációk.  
- Felhasználóbarát interfész és könnyen karbantartható kód.  
- A projekt dokumentációjának elkészítése.  
- Tesztelés és minőségbiztosítás az összes funkcióra.

## 5. Sikerkritériumok

A projekt akkor tekinthető sikeresnek, ha:  
- Az adatfeldolgozó és elemző funkciók megfelelően működnek.  
- Az eredményeket pontosan jeleníti meg és vizualizálja.  
- A kód könnyen érthető, karbantartható és bővíthető.  
- A projekt dokumentációja teljes körű és következetes.  
- A felhasználói visszajelzések pozitívak.

## 6. Kockázatok kezelése

Potenciális kockázatok:  
- Technikai nehézségek a különféle adatforrások kezelésében.  
- A fejlesztési ütemterv csúszása.  
- A szükséges modulok és csomagok elérhetőségének problémái.  
  
Kockázatkezelési tervek:  
- Alternatív adatformátumok és források beépítése.  
- Gyakori ütemezési ellenőrzések és állapotfrissítések.  
- Alternatív csomagok keresése és alkalmazása a hiányzó modulok pótlására.

## 7. Projektszervezet

Projektvezető: [Név]  
Fejlesztőcsapat:  
- Adatelemző: [Név]  
- Fejlesztő: [Név]  
- Tesztelő: [Név]

## 8. A projekt termékei

A projekt során létrejövő termékek:  
- Elemzési eredmények és vizualizációk.  
- Dokumentáció a projekt használatáról és karbantartásáról.  
- Oktatási anyagok és felhasználói kézikönyv.

## 9. Ütemterv

Durva ütemezés:  
- Adatgyűjtés és előfeldolgozás: 1-2 hét  
- Elemző és vizualizációs modul fejlesztése: 2-4 hét  
- Tesztelés és dokumentáció: 1-2 hét  
- Projekt lezárása és átadás: 1 hét

## 10. Minőségbiztosítás

A projekt minőségbiztosítási lépései:  
- Kód átnézése és optimalizálása.  
- Egységtesztek és integrációs tesztek futtatása.  
- Felhasználói visszajelzések beépítése.  
- Dokumentáció részletes felülvizsgálata.

## 11. Kommunikáció

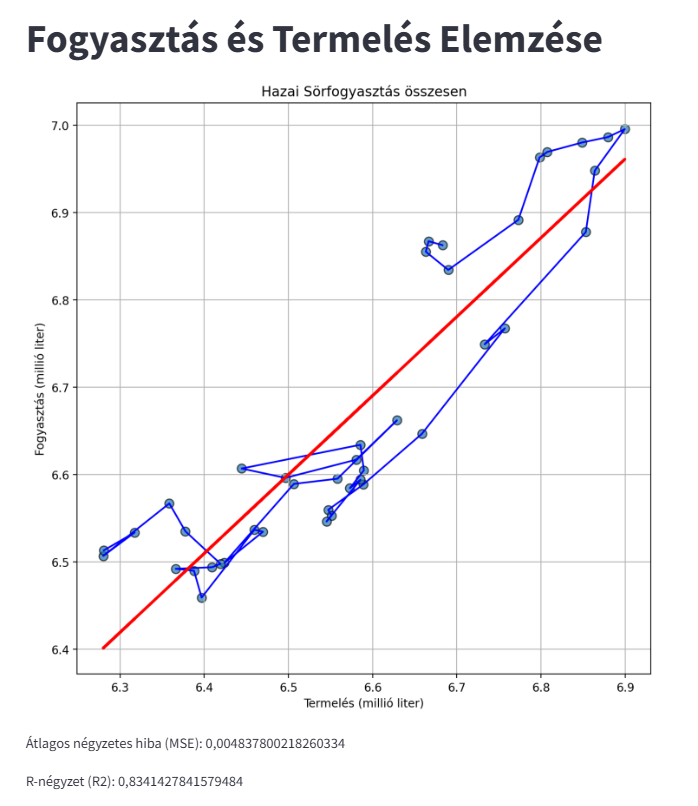
A kommunikációs csatornák és nyelvek:  
- E-mail, chat alkalmazások.  
- Dokumentáció angol és magyar nyelven.  
- Heti megbeszélések a projekt állapotának áttekintésére.

## 12. Projekt módosítások kezelése

Minden módosítást dokumentálni kell és jóvá kell hagyni a projektvezető által. A főbb változtatásokat szükséges a dokumentációban is rögzíteni.

## 13. Aláírások és mellékletek

Jóváhagyta: [Aláírás]  
Dátum: [Dátum]



A grafikon részletes elemzése az ábrázolt kapcsolat, a lineáris regressziós modell és az elemzési metrikák alapján a következő megállapításokat teszi lehetővé:

### 1. ****Grafikon Áttekintése****

* **Címek**: A grafikon címe "Fogyasztás és Termelés Elemzése", ami arra utal, hogy a cél a fogyasztás és termelés közötti kapcsolat feltárása.
* **Alcím**: "Hazai Sörfogyasztás összesen" – ez a cím egyértelműsíti, hogy a vizsgált fogyasztási adat a sörfogyasztásra vonatkozik.
* **X-tengely**: "Termelés (millió liter)" – a termelés értékeit mutatja millió liter mértékegységben.
* **Y-tengely**: "Fogyasztás (millió liter)" – a fogyasztás értékeit mutatja szintén millió literben.

### 2. ****Ábrázolt Adatok és Pontok Elhelyezkedése****

* Az egyes kék színű adatpontok a termelés és fogyasztás páros értékeit mutatják logaritmikus skálán.
* A pontok között kék vonalak vannak, amelyek az adatpontok sorrendjét mutatják, valószínűleg az időrend szerinti változást vagy az egymást követő évek termelési és fogyasztási adatait kötik össze.
* A pontok szórása meglehetősen nagy, azaz sok pont távol helyezkedik el az illesztett regressziós egyenestől, ami az adatok közötti variabilitásra utal. Ez azt mutatja, hogy bár van egy általános pozitív trend, a fogyasztás és a termelés közötti kapcsolat nem teljesen lineáris.

### 3. ****Lineáris Regressziós Egyenes****

* A piros színű vonal a lineáris regressziós modell által illesztett egyenes. Ez az egyenes reprezentálja a termelés és fogyasztás közötti becsült lineáris kapcsolatot.
* A regressziós egyenes meredeksége pozitív, ami azt jelenti, hogy a termelés növekedésével a fogyasztás is általában növekszik. Ez a pozitív trend arra utalhat, hogy a termelés és a fogyasztás között korreláció áll fenn.

### 4. ****Elemzési Metrikák****

A grafikon alatt két fontos metrikát tüntettek fel, amelyek a regressziós modell illeszkedésének minőségét mutatják:

* **Átlagos négyzetes hiba (MSE)**: 0.0048378002182603340.0048378002182603340.004837800218260334
  + Az MSE érték a valódi adatpontok és az illesztett modell által előrejelzett értékek közötti eltérést méri.
  + Minél kisebb az MSE érték, annál pontosabb az illesztés. Ebben az esetben az MSE érték viszonylag kicsi, ami azt jelenti, hogy az illesztés nem rossz, de a pontok széles szórása alapján még mindig vannak eltérések az egyeneshez képest.
* **R-négyzet (R²)**: 0.83414278415794840.83414278415794840.8341427841579484
  + Az R² érték azt mutatja meg, hogy a regressziós modell mennyire jól magyarázza az adatok variabilitását. Az értéke 0 és 1 között mozog, ahol az 1 azt jelentené, hogy az egyenes tökéletesen illeszkedik az összes adathoz.
  + Itt az R² érték 0.83, ami azt jelenti, hogy a modell az adatok varianciájának 83%-át magyarázza meg. Ez jó illeszkedésnek számít, de azt is jelzi, hogy van még egy 17%-os variabilitás, amit a modell nem magyaráz meg. Ez a fennmaradó variabilitás magyarázható lehet olyan tényezőkkel, amelyek a modellben nincsenek figyelembe véve, például szezonális hatásokkal, piaci ingadozásokkal stb.

### 5. ****További Megfigyelések és Következtetések****

* **Nemlineáris Viselkedés**: A kék pontok széles szórása és néhány adatpont jelentős eltérése a regressziós egyenestől arra utal, hogy a kapcsolat nem teljesen lineáris. Ez arra enged következtetni, hogy egy összetettebb, nemlineáris modell esetleg jobban illeszkedne az adatokra.
* **Logaritmikus Skála**: A grafikon logaritmikus skálán van, ami segíthet abban, hogy a széles skálán mozgó adatok jobban láthatóak és elemezhetők legyenek. A logaritmikus skálázás gyakran segít lineárisabb kapcsolatot feltárni az adatokban, de itt is látható, hogy néhány adatpont még mindig távol esik a lineáris egyenestől.
* **Korreláció vs. Kausalitás**: Az R² és az MSE értékek alapján kijelenthető, hogy van egy erős korreláció a termelés és a fogyasztás között, de ez nem feltétlenül jelenti azt, hogy az egyik okozza a másikat. Lehetséges, hogy mindkettőre egy harmadik, külső tényező (például gazdasági növekedés, piaci trendek) van hatással.

### Összefoglalás

* A lineáris modell jól megmagyarázza a termelés és a fogyasztás közötti kapcsolatot, de nem teljesen pontos, amit az R² érték (0.83) és az MSE is tükröz.
* Az adatok szórása arra utal, hogy egy bonyolultabb modell, például polinomiális vagy nemlineáris regresszió, esetleg jobb illeszkedést nyújthatna.
* A logaritmikus skálázás segített a vizualizációban, de a további elemzések során érdemes lehet alternatív skálákat vagy transzformációkat alkalmazni az adatok még pontosabb értelmezése érdekében.

Összességében a grafikon és a regressziós modell használható a termelés és fogyasztás kapcsolatának elemzésére, de a további finomítás érdekében érdemes lenne más modellalternatívákat is kipróbálni.