

Посвещение

С дълбока признателност посвещавам тази книга на моя изключителен ръководител, Красимир Илиев. Той не просто ми подаде ръка, а ми отвори вратата към необятния и вълнуващ свят на данните, давайки ми ценния шанс да работя като администратор на база данни.

Без неговата прозорливост, критичност и вяра в моя потенциал, нито един ред от тези страници нямаше да бъде написан. Искрено благодаря за подкрепата, напътствията и безрезервното доверие, което ми гласува. Неговата лидерска мъдрост беше и продължава да бъде моят най-силен стимул.

Тази книга е свидетелство за възможностите, които се разкриват, когато имаш ментор като него.

Благодаря Ви, г-н Илиев, от цялото си сърце!

Въведение

Добре дошли в завладяващия свят на Pandas – мощната и гъвкава библиотека на Python, която се превърна в крайъгълен камък за всеки, който се занимава с анализ и обработка на данни. Независимо дали сте начинаещ анализатор, опитен учен по данни или софтуерен инженер, който иска да разшири своя инструментариум, тази книга ще ви предостави солидна основа за ефективна работа с данни, използвайки Pandas.

В съвременния свят, изобилстващ от информация, способността да се организират, почистват, трансформират и анализират данни е все по-ценна. Pandas предлага интуитивни структури от данни и богат набор от функции, които значително улесняват тези процеси, превръщайки сложните задачи в елегантни и ефективни решения.

Този книга е структурирана така, че да ви преведе стъпка по стъпка през основните концепции и инструменти на Pandas. Ще започнем с фундаменталните типове данни и структури в Python, ще се потопим в създаването и манипулирането на Series и DataFrame – гръбнака на библиотеката. Ще разгледаме как да инспектирате вашите данни, да ги индексирате и селектирате по различни начини, да се справяте с липсващи стойности и да извършвате основни операции и трансформации. Кулминацията на тази първа част ще бъде изследването на мощните възможности за групиране и агрегиране на данни, които позволяват извличането на ценни прозрения от сложни набори от данни.

Важно е да отбележим, че тази книга предполага наличието на базови познания по програмиране и поне един съвременен програмен език. Въпреки че ще разгледаме основни концепции на Python, фокусът ни ще бъде върху специфичните инструменти и техники, които Pandas предоставя за работа с данни. Разбирането на основни програмни конструкции като променливи, цикли, условни оператори и функции ще ви помогне да усвоите по-лесно материала и да се възползвате максимално от предоставените примери и обяснения.

След като успешно преминете през тази първа част, вие ще бъдете подготвени да се потопите в по-специфични и напреднали сценарии за обработка и анализ на данни, които ще бъдат разгледани във втората част на тази книга – "Pandas за напреднали".

Пригответе се да отключите потенциала на вашите данни с Pandas! Нека заедно да направим първите стъпки в това вълнуващо пътешествие.

Кратко резюме на всяка глава

Въведение в Python за Анализ на Данни: Тази глава запознава читателя с основите на езика Python, неговите отличителни черти и защо е предпочитан за анализ на данни. Разглеждат се основни типове данни, управляващи структури, функции, основи на обектно-ориентирано програмиране, обработка на изключения, работа с файлове и модули, както и въведение в библиотеката NumPy и средите за разработка.

Основи на Pandas: Тази глава представя библиотеката Pandas, нейните основни предимства и приложения в анализа на данни. Въвежда основните структури от данни в Pandas – Series и DataFrame, обяснява как да ги създаваме от различни източници и разглежда основните им атрибути за получаване на информация.

Основно Разглеждане и Инспекция на Данни: Тази глава се фокусира върху първите стъпки при работа с DataFrame-и, като показва как да преглеждаме първите и последните редове, да получаваме обща информация за структурата и типовете данни, да изчисляваме описателна статистика и да изследваме уникалните стойности.

Индексиране и Селектиране на Данни: Тази глава е посветена на начините за достъп и извличане на конкретни данни от Series и DataFrame. Разглеждат се различни методи за селектиране по колони и редове, използвайки скоби, `.loc`, `.iloc`, булево индексиране и други техники за филтриране и извличане на подмножества от данни.

Индексни Обекти (Index Objects): Тази глава разглежда ролята и значението на индексите в Pandas, които осигуряват етикетиране и подравняване на данните. Представят се различните типове индексни обекти, които Pandas предлага за работа с числови, времеви, категорийни и други видове данни.

Обработка на Липсващи Данни: Тази глава е посветена на проблема с липсващите данни в наборите от данни. Разглеждат се начини за идентифициране, преброяване, премахване и запълване на липсващи стойности, както и техники за интерполиране.

Операции и Манипулации на Данни: Тази глава обхваща основните операции за трансформиране и манипулиране на данни в Pandas. Разглеждат се векторизирани операции, прилагане на функции (елемент по елемент и по оси), създаване и преименуване на колони и индекси, сортиране и промяна на типовете данни, както и работа с текстови и времеви данни.

Комбиниране и Сливане на DataFrame-и: Тази глава се фокусира върху техниките за комбиниране на данни от различни DataFrame-и. Разглеждат се конкатенация за обединяване по редове или колони и сливане (`merge`) в стил бази данни за свързване на DataFrame-и въз основа на общи колони или индекси.

Групиране и Агрегиране (GroupBy): Тази глава представя мощния механизъм на Pandas за групиране на данни въз основа на определени критерии и прилагане на агрегиращи функции за обобщаване на информацията във всяка група. Разглеждат се методи за инспекция на групи, итериране, трансформация и филтриране на групи, както и гъвкавото прилагане на произволни функции.

Зареждане и Записване на Данни (Input/Output): Тази глава е посветена на начините за четене на данни от различни файлови формати (CSV, Excel, JSON, HTML, SQL и др.) и записване на обработените данни обратно във файлове, което е ключова стъпка в процеса на анализ на данни.